

STIC-Biotech/Ch mLib

76159

From: Rao, Manjunath N.  
Sent: Friday, September 20, 2002 10:12 AM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 09/836613

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10C 01  
Phone: 306-5681

Date: 9-20-02

RECEIVED  
SEP 20 2002  
STIC

Please search the following as soon as possible for application with serial number **09/836,613**

**SEQ ID NO:2**, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If y u have any questions please call me at the above phone number.

Thanks

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

Manjunath N. Rao, Ph.D.  
Biot chnology Patent Examiner  
Art Unit 1652, Room 10A11  
Crystal Mall 1, USPTO.

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 9/21/02  
Date Completed: 9/21/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: ✓  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2002, 12:23:45 ; Search time 34.74 Seconds  
(without alignments)  
2375.585 Million cell updates/sec

Title: US-09-836-613-2  
Perfect score: 3939  
Sequence: 1 MEAVAVAAAGVLLLAGAGG.....VDLAKKIFLKYYPGWAGSW 743

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1993.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1989.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1990.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1991.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1992.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1993.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1994.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1995.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1996.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1997.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1998.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA1999.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA2000.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID          | Description         |
|------------|-------|-------------|--------|-------------|---------------------|
| 1          | 3939  | 100.0       | 743    | AAW18017    | Human alpha-N-acet  |
| 2          | 1407  | 35.7        | 690    | 22 ABB68119 | Drosophila melanog  |
| 3          | 125.5 | 3.2         | 497    | 22 AAE05171 | Human drug metabol  |
| 4          | 121   | 3.1         | 4342   | 22 AAU33611 | Pseudomonas aerugi  |
| 5          | 121   | 3.1         | 4472   | 18 AAU22601 | Tyrlactone synthase |
| 6          | 120   | 3.0         | 3729   | 18 AAU22603 | Amino acid sequenc  |
| 7          | 118.5 | 3.0         | 871    | 22 AAB30570 | A full length huma  |
| 8          | 118.5 | 3.0         | 1351   | 22 AAB30567 | Human protein kina  |
| 9          | 118.5 | 3.0         | 1610   | 22 AAB30568 | A splice variant o  |
| 10         | 118.5 | 3.0         | 1618   | 22 AAB85504 | Human protein kina  |
| 11         | 118.5 | 3.0         | 2596   | 22 AAB30569 | A splice variant o  |

|    |       |     |      |    |          |                    |
|----|-------|-----|------|----|----------|--------------------|
| 12 | 117.5 | 3.0 | 497  | 16 | AAU72376 | Human auxillary cy |
| 13 | 117.5 | 3.0 | 497  | 16 | AAU72377 | Human auxillary cy |
| 14 | 117.5 | 3.0 | 497  | 17 | AAU93183 | Human cytochrome p |
| 15 | 117.5 | 3.0 | 497  | 17 | AAU93184 | Human cytochrome p |
| 16 | 117.5 | 3.0 | 497  | 17 | AAU81462 | Human derived cyto |
| 17 | 117.5 | 3.0 | 497  | 19 | AAU44869 | Cytochrome P450Iid |
| 18 | 113   | 2.9 | 7257 | 21 | AAU58576 | Sorangium cellulos |
| 19 | 111.5 | 2.8 | 484  | 21 | AAU58576 | Arabidopsis thalia |
| 20 | 111.5 | 2.8 | 497  | 16 | AAU72375 | Human auxillary cy |
| 21 | 111.5 | 2.8 | 497  | 17 | AAU72378 | Human auxillary cy |
| 22 | 111.5 | 2.8 | 497  | 17 | AAU93182 | Human cytochrome p |
| 23 | 111.5 | 2.8 | 497  | 17 | AAU93185 | Human cytochrome p |
| 24 | 109.5 | 2.8 | 1377 | 22 | AAU34791 | E. coli cellular p |
| 25 | 109.5 | 2.8 | 1377 | 22 | AAU98997 | E. coli growth and |
| 26 | 107.5 | 2.7 | 1621 | 19 | AAU60182 | Arabidopsis thalia |
| 27 | 106   | 2.7 | 308  | 22 | AAU92974 | C glutamicum prote |
| 28 | 106   | 2.7 | 1005 | 22 | ABG11447 | Novel human diagno |
| 29 | 105   | 2.7 | 1203 | 20 | AAU81572 | Mus dunni endogeno |
| 30 | 105   | 2.7 | 4150 | 21 | AAU92707 | S. antibioticus 8, |
| 31 | 104.5 | 2.7 | 1622 | 19 | AAU60163 | Arabidopsis thalia |
| 32 | 104   | 2.6 | 581  | 21 | AAU84322 | A human cardiovasc |
| 33 | 104   | 2.6 | 581  | 22 | AAU65701 | Novel protein kina |
| 34 | 104   | 2.6 | 588  | 21 | AAU58161 | Adeno associated v |
| 35 | 104   | 2.6 | 654  | 22 | AAU29036 | Human PRO polypept |
| 36 | 104   | 2.6 | 654  | 22 | AAU06593 | Human protein havi |
| 37 | 104   | 2.6 | 654  | 22 | AAU80243 | Human PRO282 prote |
| 38 | 104   | 2.6 | 668  | 22 | AAU67462 | Amino acid sequenc |
| 39 | 104   | 2.6 | 677  | 17 | AAU99539 | Soluble starch syn |
| 40 | 104   | 2.6 | 724  | 21 | AAU58160 | HIV-1 gp160-NV5 p5 |
| 41 | 103   | 2.6 | 1188 | 17 | AAU94622 | Adeno associated v |
| 42 | 102   | 2.6 | 1646 | 22 | AAU79494 | Human protein SEQ  |
| 43 | 102   | 2.6 | 2308 | 15 | AAU57902 | Human RPTP-beta.   |
| 44 | 101.5 | 2.6 | 5069 | 19 | AAU52846 | A. mediterranei ri |
| 45 | 101   | 2.6 | 1231 | 22 | ABG17594 | Novel human diagno |

## ALIGNMENTS

RESULT 1  
AAW18017  
ID AAW18017 standard; Protein; 743 AA.  
XX AAW18017;

20- AUG-1997 (first entry)

XX Human alpha-N-acetylglucosaminidase.

DE Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;  
KW gene therapy; enzyme replacement therapy; diagnosis.  
XX Homo sapiens.

| Key              | Location/Qualifiers                     |
|------------------|---|
| FT Peptide       | 1..23                                   |
| FT Protein       | /label= Sig_peptide                     |
| FT Protein       | 22..743                                 |
| FT Protein       | /label= Mat_protein                     |
| FT Modified-site | 261                                     |
| FT Modified-site | /note= "potential N-glycosylation site" |
| FT Modified-site | 272                                     |
| FT Modified-site | /note= "potential N-glycosylation site" |
| FT Modified-site | 435                                     |
| FT Modified-site | /note= "potential N-glycosylation site" |
| FT Modified-site | 503                                     |
| FT Modified-site | /note= "potential N-glycosylation site" |
| FT Modified-site | 513                                     |
| FT Modified-site | /note= "potential N-glycosylation site" |
| FT Modified-site | 526                                     |
| FT Modified-site | /note= "potential N-glycosylation site" |
| FT Modified-site | 532                                     |
| FT Modified-site | /note= "potential N-glycosylation site" |

XX W09719177-A1.  
PN  
XX PD  
XX  
XX 29-MAY-1997.  
XX  
XX 22-NOV-1996; 96WO-AU00747.  
XX PF  
XX 23-NOV-1995; 95AU-0006748.  
XX PR  
XX (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.  
XX PA  
XX  
XX Anson DS, Blanch L, Hopwood JJ, Scott H, Weber B;  
PI WPI; 1997-298114/27.  
XX  
XX N-PSDB; AAT67163-64.  
XX  
XX Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -  
PT used for the diagnosis and treatment of mucopolysaccharidosis type  
PT IIIB, also used in gene therapy  
PT  
XX  
XX Claim 11; Page 46-50; 79pp; English.  
XX  
XX Human alpha-N-acetylglucosaminidase (AAW18017) can hydrolyse the  
CC terminal alpha-N-acetylglucosamine residues at the non-reducing  
CC terminus of fragments of heparan sulphate and heparin. Its amino  
CC acid sequence was deduced from a human peripheral blood leukocyte  
CC cDNA clone (AAT67163). Recombinant enzyme can be expressed in host  
CC (pref. CHO) cells and may be expressed as a fusion to e.g. an  
CC enzyme, reporter molecule, purification tag and/or signal sequence.  
CC It can be used to treat alpha-N-acetylglucosaminidase deficiency.  
CC for example in patients suffering from mucopolysaccharidosis type  
CC IIIB. Administration is by oral, i.v., i.p., enzyme replacement  
CC therapy, gene therapy or other routes.  
XX  
XX Sequence 743 AA;  
SQ

Query Match 100.0%; Score 3939; DB 18; Length 743;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAVAAVAAVGVLLLAGAGGAGDEAREANAARALVARLLGCPAADEFSSVERALAAKP 60  
DB 1 meavavaavgvlllagaggagdeareanaavralvarllgpggaadfsvsveralaakp 60

QY 61 GLDITYSLGGGAARVRVVGSTGVAAAAGLHRYLRDFCGCHVAWSGSQLRPLRPAPVQEE 120  
DB 61 gldtyslsgggaarvrvgstgvaaaaglhrylrdfcgchvawsgqlrlprlpavpge 120

QY 121 LPEATPNRYRYQNVCTQSYSTVWWDWARWEREIDWMALNGINLALAWSGQEAIRQRYVL 180  
DB 121 lpeatpnryryqnvctqsystvwwdwarwereldwmalinglnlalswsgqeaigrvyl 180

QY 181 ALGLTQAEINEFTTQAFPAWGRMGLHTWDCGPLPSPWHIKOLYLQHRVLDQMRSGFGWTP 240  
DB 181 alglqtqaeinefttqafpawgrmglhtwdcgplpsspwhiklylqhrvldqmrsgfgmt 240

QY 241 VLPFAGHVPFAVTRVFPQVNTKMGSMGFNCSSCSFLLAPEDPFIPIGSLFLRELI 300  
DB 241 vlpfaghvpeavtrvfpqvnvtnkmgswghfncsscsfllapedpfpilgslflreli 300

QY 301 KEFGTDHYGADTFNEMQPPSPSEPSYLAATTAATVAVTEAVTLQGLHGFQHPQF 360  
DB 301 kefgtdhygadtfnemqppspsepsylaattaatvavteavtlqglhgfqhqpqf 360

QY 361 WCPAQITRAVLGAPVGRLLVLDLFAESQPVYTRTASFOQPTIWCMLHNFNGHGLFGAL 420  
DB 361 wcpaqitravlgapvgrllvldlfaesqpvvtrtasfgqpfiwcmhlnfgnghlfgal 420

QY 421 EAVNGGPEAARLPFNSTWGTGMAPEGISQNEVYVSLMAELGWRKDPVPDLAAWVTSFAA 480  
DB 421 eavnggpeaarlpfnstwgtgmapegisqnevvyslmaelgwrkdpvpdlaaawvtsfaa 480

QY 481 RRYGVSHPDAGAAWRLLRSVYNCSGEACRGHNRSPLYRRPSLQMNNTSIWYNRSDVFEAW 540  
DB 481 rrygvshpdagaawrllrsvynsgsgeacrghnrsplvrrpslqmntsiwynrsdvfeaw 540

QY 541 RLLTSPATSPAFRYDLDLTRAQVQELVLYVEEARSAVLSKELASLLRAGGVLAY 600  
DB 541 rlltspatspafrydldltravqelvlyveearsavlskelasllraggvlay 600

QY 601 ELLPALDEVLASDRSFLGSLWLEQARAAVSEAEADFYEQNSRYQLTLWGPENGLIDYAN 660  
DB 601 ellpaldevlasdsrflgslwleqaraaavseaeadfyeqnsryqltlwgpennldyan 660

QY 661 KQLAGLVANYTTPRWRLFEALYDSVAQGIFFQHQHFDKNVFOLEQAFVLSKQRYVPSOPR 720  
DB 661 kqlaglvanytprwrlfealvdsvaqgiFFQHQHFDKNVFOLEQAFVLSKQRYVPSOPR 720

QY 721 GDTVDLAKKIFLKYYPGWVAGSW 743  
DB 721 gdtvolakkiflkyypgwvagsw 743

## RESULT 2

ABB68119  
ID ABB68119 standard; Protein; 690 AA.  
XX  
AC ABB68119;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 31149.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL12222.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 31149; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 690 AA;

```
Query Match      35.7%:  Score 1407;  DB 22;  Length 690;
Best Local Similarity 43.3%:  Pred. No. 1.4e-118;
Matches 290;  Conservative 102;  Mismatches 255;  Indels 22;  Gaps 11;

QY 74 RVRVGGTGVAAAAGLHRYLRDFCCCHVANGSGQLRLPR--PLPAVPGELTEATPNRYR 131
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 6 fillmgwgvsvckalhhlykyvinkdvkfmkleiptnlqpnvtiesksasp--ily 63
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 132 YQNVCTOSYFVWMDWREIDWALNGINLALAWSGOEAIWQRYLALGLTQAEINE 191
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 64 hqnvctysfawgieqrhrhdwalmgisitia-pvqea1wkvtydmgrmeide 122
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 192 FFTGPAFLANGRGNLHTWDPGLPPSHIHKOLYLQHRVLDQMSFGMTPLPAFAGHVE 251
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 123 hlagpatgawgrnirgawgipitpawryqlllqqeiltacnlgmsvalpafaghvpr 182
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 252 AVTRVFQVNVTKMGSMGHFNCSYCSFLAPEDPPIPIIGSLFLRELKIEFGTTHYGA 311
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 183 alkrlnpestfmevqrwnqfpdryccglfveptenlfkelastrflhniitkygsnhffc 242
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 312 DTNEMOPSPSEPSYLAATTAAYEAMTAVDTEAVYVLLQGLWFOHQFQWGPQAQIRAVLG 371
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 243 dpineleppvakpeymrstaaiyesmrgidpqaiwlgwmfvknp-fwtldmaeafit 301
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 372 AVPRGRLVLDLFAESQPVYTRTASQGPFTWCMHNFNGHGLFGALEAVYNGGPEAAR 431
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 302 aapgrilvldlqsedfpdyeltrsyfgqpficwlmhmfgtlmgfsgsaklinsgleear 361
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 432 LFPNSTMVGTMAGPEGISNEVYSLMAELGWRKDPVPLAAWVTSFAARRYGVSHPDAG 491
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 362 rlpnssilvgitpegigqymysftlrgwsntsi-dldswftnfishrygvkderle 420
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 492 AAWRLLLRSVYNCSC--EACRGNNRSLVRRPSLQMNSTSIWYNRSDVFAWRLLLTSAPSL 550
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 421 qawlliknsyfrgikmrg--qyvvtirpsfngqepftwynaasavidawhlllfrail 478
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 551 ATS----PAFRYDLDLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLAYELLPAL 606
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 479 p1ednryelghdvdtitqfiasdqlylnrsaytkrqvsrf----eflsvklklf 534
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 607 DE---VLASDRFLGLSMLEQARAAVSEAEADFYEQNSRVLTLWPGNGLDYANKOL 663
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 535 dmeillassrnlflgnwlqqakapntgqgrnfefnarnqitawpgdgildyackqw 594
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 664 ACLVANYTPRWRLFEALVDSVAQIPFQHQFDKNV-FOLEQAFVLSKORYPSQPRGD 722
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 595 sglvsdyrprwrlfledvtvalhagrpngtatk1kvshelelpfsnknkdvyvptpvg 654
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 723 TVDLAKKIF 731
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 635 twllsqdif 663
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
AAE05171
XX AAE05171 standard; Protein; 497 AA.
XX AC AAE05171;
XX DT 12-SEP-2001 (first entry)
XX DE Human drug metabolising enzyme (DME-2) protein.
XX KW Human; drug metabolising enzyme; DME-2; immunosuppressive; gene therapy;
KW KW cystostatic; autoimmune disorder; inflammatory disorder; atherosclerosis;
KW KW osteoporosis; eye disorder; hepatic tumour; Addison's disease; cretinism;
KW KW rheumatoid arthritis; acquired immune deficiency syndrome; AIDS; anaemia;
KW KW developmental disorder; endocrine disorder; iritis; acromegaly; epilepsy;
KW KW thyrotoxicosis; pancreatic disorder; diabetes mellitus; obesity; adenoma;
KW KW gastrointestinal disorder; nodular hyperplasia; conjunctivitis; glaucoma;
KW KW actinic keratosis; metabolic disorder; dysphagia; anorexia; carcinoma;
KW KW cell proliferative disorder.
XX
```

Homo sapiens.

Key Peptide Location/Qualifiers  
1..25 /label= Signal\_peptide  
26..497 /note= "Mature drug metabolising enzyme (DME-2) protein"  
1..18 /label= Transmembrane\_domain

WO200151638-A2.

19-JUL-2001.

12-JAN-2001: 2001WO-US01174.

14-JAN-2000: 2000US-0176139.

21-JAN-2000: 2000US-0177443.

28-JAN-2000: 2000US-0178574.

(INCY-) INCYTE GENOMICS INC.

Yang J, Baughn MR, Burford N, Au-Young J, Lu DAM, Reddy R;  
Ring HZ, Hillman JL, Yue H, Azimzai Y, Yao MG, Gandhi AR;  
Nguyen DB, Tang YT, Lal P, Bandman O;

WPI: 2001-425874/45.  
N-PSDB: AAD09937.

Drug metabolizing enzymes and encoding polynucleotides, useful for  
diagnosing, treating and/or preventing autoimmune, inflammatory, cell  
proliferative, developmental, endocrine, eye, metabolic, and  
gastrointestinal disorders -

Claim 1; Page 136-137; 133pp; English.

The present sequence is human drug metabolising enzyme (DME-2) protein.  
Human DME and its nucleic acid molecule are useful for the diagnosis,  
treatment and prevention of disorders associated with increased or  
decreased expression of DME. Examples of such disorders include,  
autoimmune/inflammatory disorder such as acquired immune deficiency  
syndrome (AIDS), rheumatoid arthritis, osteoporosis; cell proliferative  
disorder such as actinic keratosis, atherosclerosis; developmental  
disorder such as epilepsy, anaemia; endocrine disorder such as  
acromegaly, cretinism, thyrotoxicosis; pancreatic disorder such as  
diabetes mellitus; eye disorder such as conjunctivitis, glaucoma, iritis;  
metabolic disorder such as Addison's disease, obesity; gastrointestinal  
disorder such as anorexia, dysphagia and hepatic tumours including  
nodular hyperplasia, adenomas and carcinomas. DME DNA is useful for  
creating 'knockin' humanised animals (pigs) or transgenic animals (mice  
or rats) to model human disease. DME DNA is also useful for gene  
therapy. DME and its immunogenic fragments are useful for screening  
libraries of compounds in several drug screening assays.

Sequence 497 AA;

Query Match 3.2%: Score 125.5; DB 22; Length 497;  
Best Local Similarity 21.9%; Pred. No. 0.023;  
Matches 93; Conservative 53; Mismatches 143; Indels 135; Gaps 21;

QY 345 AVWLL-----QGWLFQHFQFQWGPQAQIRAVLGAVPRGRLVLDLFAESQPVYTRTAS 396  
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 14 afillvldmhrqrwaayrpp---gplplpgl-----gnllhvd----- 50  
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 397 FQCGQPIWCMHNFNGN-----HGLFGALEA-VNGGPEAARL--FPNSTMV 439  
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 51 fntpycfqdlrrrrgdfvslglawtpvvlnglaavrealvthgedtadrpvpitqil 110  
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 440 GTGMAPEGI-----SQNEVYSLMAELGWRKDPVPLAAWVTSFAARRYGVSHPD 489  
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 111 gfgsrsqgvflarypawreqrfsvstirnlgk---ksleqwteteeaa----- 158  
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 490 AGAARLLLSVYNGSGACGRNRSPLVRPSLOMTNYSIWNRSYVEANRLLTSAPS 549  
Db 159 -----CLcaafanhgrpf--rpnglldkav-----snv1-----as 188  
QY 550 LATSPAFRYD-----LIDLROAVOELVSLVYEARSAYLSKELASLLRAGGVLAYE-- 601  
Db 189 ltcgrfeydprflrldldaqeikesgfirevlnavpvlphpal--agkvlrfoka 246  
QY 602 LLPALDEVL-----ASDSRFLGSLWLEQARAAVSEAEADYEQNSRYQLTLWGPE 652  
Db 247 fltqdeltehrmtwpaqpprdlteafak-kekagsspsfndenlrivv----- 299  
QY 653 GNILDYANKOLAGLVANYTPWRFLFLEALVDSVAGGTFQOHQFDKNVQLEQAFVLSK 712  
Db 300 gnlf-----lagmvtsttllawallmlhpdvqcrv---qqeidevigvrhpevadq 350  
QY 713 QRPV 716  
Db 351 ahmp 354

RESULT 4  
AAU33611  
ID AAU33611 standard; Protein: 4342 AA.  
XX AAU33611;  
AC AAU33611;  
XX  
DT 14-FEB-2002 (first entry)  
DE Pseudomonas aeruginosa cellular proliferation protein #55.  
XX  
DE Antisense: prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.  
KW  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
DR WPI; 2001-611495/70.  
DR N-PSDB; AAS51470.  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
PS Example 3; Seq ID NO 5107; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 4342 AA;

Query Match 3.1%; Score 121; DB 22; Length 4342;  
Best Local Similarity 21.4%; Pred. No. 1.8;  
Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;

QY 4 VAVAA-----AVGVLLLAGAGG-----AGDEAREAAAVRA-LVARLL 40  
Db 1181 vaicaerspqlivgllaivkagayvpdpdpaperlaymladsgvelllctqahlferl- 1239  
QY 41 GPG-----PAADFSVSEVERALAAKPGDLTYSLGGGAARVRVSGTGVAAAGLHRYLRD 95  
Db 1240 -pgaegvtpicldskldnwpqapglhlh--gdnlayviytsgstgqpkvgv----- 1289  
QY 96 FCGCHVAMSGSOLRLRPLPAVPGE--LTEATPNRYRYQNVCTOSYSFVWMDWARWERE 153  
Db 1290 ---nthaaala-erlqwmqatytldgddvImqkap-----vfdvsvw--e 1328  
QY 154 IDWMALNGINLALAWSGQ-----EAIQWRVYLALGLTQAEINEFFTPGPAFLAWGRMG 205  
Db 1329 cfwplvtgcrvllaapgehrdparivelvrgfvtllhtfvpplqlqfidepgvaacgsair 1388  
QY 206 NLHTWDGGLPPSNHKLQYLQHRVLDQMRSGFTVLPFAFAGHVPEAVTRVFPVQNVTKM 265  
Db 1389 rlfsggealpa-----lrnrvlqr-----lpavalhnrvgptet--ainvt-- 1428  
QY 266 GSWGHEFNCYSFLLAPED---PIFPIIGSLFLRELKEFGTDHIYGADTFNEM--- 317  
Db 1429 ----hwgc-----raedgersplgrplgnvvcvldaef---nllpagvageicgg 1473  
QY 318 -----QPPSEPSYLAATTAAYEAMTAVDTEAVWLLQGLW-----FQHQQFQWG- 362  
Db 1474 lglargylgrpalsaeafvadpfsaagerlyrtgdrrwnadgvieyigrldqqvklrgf 1533  
QY 363 ---PAQIRAVLGAVP---RGRLLVLDLFAESQPV--YTRTASFQGPFFIWCMLHNFQGNH 414  
Db 1534 riepeelgarllaqpvagavviregvgsgqlvgytgvagaeaea-----eqnq 1584  
QY 415 GLFGALEAVNGGPEAARLPNSTMTVGTGMAPEGISONEVVYSMLAELGWRK-----DPVPD 470  
Db 1585 rlrhaalga-----elpeymvptqlmrlagmplggs-gkidtralpepvwqrehveprte 1638  
QY 471 L-----AAWVTSFAARRYG-----VSHPDAGAAWRLLLRSVYVNCs-- 505  
Db 1639 lqriaaiwseviglprvgirddffellghslhatrivstrqacdvplralfeasel 1698  
QY 506 -----GEACGHNRSPVRRPSLQMNSTSIWNRSDVFEAMRLLTSAPSLSATSPAF 556  
Db 1699 eafceqvraaqagrtshgaairrdreqvpplaysqgrmfwlwl-----epd---spay 1751  
QY 557 RY-----DLDLLTR--QAVQELVSLVYEARSAYLSKELASLLRA---GGV-LAYELL 603  
Db 1752 nvvgglarlsgpldvarfeaalqalvq-rhetlrrtspdvqgvprvhgdggimdwqdf 1810  
QY 604 PALD-----EVLASDRSFLGSLWLEQA---RAAAVSEAEADFYEQNSRYQLTLWGPE 652  
Db 1811 saldrdrqhlqlcla-dseahrpfldespllrvcvmkmaerehyllvltlhhiv---te 1866  
QY 653 GNILDYANKOLAGLVANYTPWRFLFLEALVDSVAGGTFQOHQFDKNVQLE 705  
Db 1867 gwamditarelgalyeaflddr-----esplelpvq--yldysvwqre 1908



XX Tyactone synthase ORF3 protein.  
DE Tyactone synthase gene cluster; tylG gene; multifunctional protein;  
KW polyketide; tyactone synthesis; antibiotic; tylosin.  
KW Streptomyces fradiae.  
OS  
XX Key Location/Qualifiers  
FH Misc-difference 1 /note= "encoded by GTC"  
FT Domain 43..465  
FT /note= "ketosynthase domain, KS4"  
FT Domain 588..916  
FT /note= "acyltransferase domain, AT4"  
FT Domain 1178..1356  
FT /note= "ketoreductase domain, KR'4"  
FT Domain 1451..1534  
FT /note= "acyl carrier protein domain, ACP4"  
FT Domain 1556..1980  
FT /note= "ketosynthase domain, KS5"  
FT Domain 2095..2425  
FT /note= "acyltransferase domain, AT5"  
FT Domain 2451..2636  
FT /note= "dehydratase domain, DH2"  
FT Domain 3274..3453  
FT /note= "ketoreductase domain, KR5"  
FT Domain 3548..3631  
FT /note= "acyl carrier protein domain, ACP5"  
XX EP79M55-A2.  
PN  
XX 27-AUG-1997.  
PD  
XX 19-FEB-1997; 97EP-0301056.  
PF  
XX 22-FEB-1996; 96US-0012078.  
PR  
XX (ELIL ) LILLY & CO ELI.  
PA  
XX Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;  
PI WPI; 1997-418046/39.  
DR N-PSDB; AAT80413.  
DR  
XX DNA encoding Streptomyces fradiae tyactone synthase domain - for  
XX production of tylosin-related polyketide compounds  
XX Claim 9; Pages 86-98; 220pp; English.  
XX  
XX AAW22601-W22605 represent proteins encoded by the tyactone synthase  
XX gene cluster of the invention. The gene cluster is also referred to as  
XX the tylG gene, and was isolated from Streptomyces fradiae. These  
XX sequences are multifunctional proteins which direct the synthesis of the  
XX polyketide tyactone, isolated from Streptomyces fradiae. Tyactone is  
XX the basic building block of the antibiotic tylosin. The DNA sequence can  
XX be modified so as to alter the type of carboxylic acids incorporated.  
XX The number of carboxylic acids incorporated and/or the post-condensation  
XX reactions performed, thereby resulting in novel tylosin-related  
XX polyketides.  
XX  
XX Sequence 3729 AA;  
SQ  
Query Match 3.0%; Score 120; DB 18; Length 3729;  
Best Local Similarity 22.4%; Pred. No. 1.7;  
Matches 192; Conservative 72; Mismatches 265; Indels 328; Gaps 47;  
QY 4 VAAVAARV---GVLLLAGAGGAGCDEAREAAVRLVRLIG-----PGPAAD 47  
DB 668 vdaavvghsgelaaatvagalsle---daavvalragligrlylagamaavplpage 725  
QY 48 FSVSVERALAAKPGDLTYSLGGGAA-----RVRVRGSGTGVAAAAGLH----- 90

DB 726 ---veaglakwpgveavaavngpastvsgdrravagvavcqaegvarlipdyashs 781  
QY 91 RYLRDFCG-CHVANGSGLRLPR-PLPAV-----PGLTEATPNRYRYQNVCTQSYSFV 143  
DB 782 rhvedlkgelervlsglrprsvrvcstvageqgepv----- 820  
QY 144 WMDWARWER-----EIDMMALNGINL-----ALAWSGQPAIWQRYLA 181  
DB 821 -fdagywfrlnrrvefsavvggllleghrrfievhsahpvlvhaieqtasea--dravha 877  
QY 182 LGLTQAE---INERTGPAPFLAWGRMGNLHTWDGPPSWHIKOLYL-----QHRVLDQM 233  
DB 878 tglrrqgdsphrltsta-eaahgatl-twdpalppg-hittlptypfhhhhywid-- 932  
QY 234 RSFGMTPVLPAPF-----AGHVPEAVT--RVFPQVNVTKM--GSWGHFNCSYSCS 278  
DB 933 ---tpttptattgspcdagpnadapkykvwkrlrdqdsaltarldgrw----- 978  
QY 279 FLAPEDPIFFIIGSLFLRELK-----EFGTDH-----IYGADTFNEMQPPSS 322  
DB 979 llivpeasadspsaegvareitargetvesltvepgadrsrlrgllvdaterdeagplrg 1038  
QY 323 EPSYLA-----AATTAVYEAMTAVDTEA-VMLQGLWLFQHQPQFWG 362  
DB 1039 lvsllalagdhagadgarvvpaglaasialliqaagdagteeglw----- 1083  
QY 363 PAQIRAVLGAVPRGRLLVLDLFAESQPV---YTRTASQGGQFFIWCMLHNFNGNHLGFG- 418  
DB 1084 -avtvgavaavpg-----dvpapsqallwgfgrvagie-lphcwggldldtptgpdgfg 1135  
QY 419 -----ALEAVNGPEEARLPNSTMVGT-GMAPEGISONEVVYSIMAE 461  
DB 1136 rqlaatlagrpaedqvalra--sgayrrflvrasaaggadgwrprg-----tlvv 1184  
QY 462 GWRKDPVPDLAAWVTSFAARRY---GVSHP-----DAGAAWRLLLRSVYNCSEACRGHNR 514  
DB 1185 gdtavagplvrlwlgngarrvtlsgplpeeladvaarvtv-----apcdpadr 1236  
QY 515 SPLVRRPSLOMNTSIWNRSDVFEAWRLLLTSAPSLATSPAFRYDLDLTROA----- 567  
DB 1237 palrtllaegaptav-----lvappavppcp-----Jaemtaealalalaea 1277  
QY 568 ---VQELVSLVYE-----EARSAYLSKELASLL 592  
DB 1278 ktglvrlsdlldpplledgeldafvfvssvavvgggaggvaggayagtayid-alaecr 1336  
QY 593 RAGG--VLAYELLPALDEVLASDSRFLLSWLEQA-----RAAAV 630  
DB 1337 ragglpvtsvawtpwig-tpaads---lgeqmsragltpldpaasldalaravgrragcv 1392  
QY 631 SEAEADFYEQNSRYQLT 647  
DB 1393 tvadidwerfasaytat 1409  
RESULT 7  
AAB30570  
ID AAB30570 standard; Protein; 871 AA.  
XX  
AC AAB30570;  
XX  
XX 19-MAR-2001 (first entry)  
DT  
XX  
XX A splice variant of a signal transduction polypeptide.  
DE  
XX Signal transduction; H19C5; kinase; cardiac disease; angina pectoris;  
KW congestive heart failure; dilated congestive cardiomyopathy;  
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
KW atherosclerosis; cardiac tumour; microbial infection; splice variant.  
XX



OS Homo sapiens.  
 PN WO200063381-A1.  
 XX 26-OCT-2000.  
 XX 11-APR-2000; 2000WO-US09488.  
 XX 16-APR-1999; 99US-0129553.  
 XX (SCIO-) SCIOS INC.  
 XX Zeng W, Stanton L, Kong H;  
 XX WPI; 2001-007013/01.  
 XX Novel h19G5 polypeptides capable of regulating signal transduction and  
 PT exhibiting kinase activity useful for identifying antibodies to treat  
 PT cardiac diseases, and additional mediators of signal transduction -  
 XX  
 PS Claim 1; Page 74-76; 81pp; English.  
 XX  
 CC The present sequence represents a splice variant of human in signal  
 CC transduction polypeptide. The polypeptide is designated H19G5. The  
 CC protein is capable of regulating signal transduction and exhibits kinase  
 CC activity. The H19G5 transcript is expressed in the heart. H19G5  
 CC polypeptides and polynucleotides are useful for preventing or treating a  
 CC cardiac disease, such as congestive heart failure, dilated congestive  
 CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,  
 CC mitral valve disease, aortic valve disease or tricuspid valve disease,  
 CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,  
 CC arterial or rheovascular hypertension, arteriosclerosis, atherosclerosis  
 CC and cardiac tumours in humans. The polypeptide is also useful for  
 CC detecting the expression of a protein capable of regulating signal  
 CC transduction or the expression of a phosphate group. The monoclonal antibodies can  
 CC or acceptor molecule of a phosphate group. The monoclonal antibodies can  
 CC be used as probes for detecting discrete antigens expressed by tissue or  
 CC cell samples, and therefore used in humans for localization and  
 CC monitoring of microbial infection.  
 XX  
 SQ Sequence 871 AA;

Query Match 3.0%; Score 118.5; DB 22; Length 871;  
 Best Local Similarity 18.6%; Pred. No. 0.24;  
 Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

Qy 19 GGAAGDEAREAAVALVARL-LGPGPAAD---FSVSVERALAAKPGLD-----TY 65  
 Db 21 gsasqassqvsslrvgssqvgtepgpsldaegtqeadlsdstptlqrpqeqatmrkf 80  
 Qy 66 SLGG-GGAARVRVRCSTGVAAGLHRYLRDFCG-----CHVAMSGSGLR----- 109  
 Db 81 sl9rggyagvagygtfafgagdag-----gmlgggpmwariawavsgseeeqearae 134  
 Qy 110 -----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSYSPVWMDWARWE 151  
 Db 135 sqseeqgearaeslpqvsarvpvevgrptrssp-----eptpwe 175  
 Qy 152 -----REI---DWMALNGINLAL-----AWSGOEAIWORYLAL----- 182  
 Db 176 dlqgvslvqirdlsgdaaatisidisevdpaylnlsldydikylpfefmfrkvpksa 235  
 Qy 183 -----GLTQAEINEF-----FTGP-----AFLANGRMGNLHTWDGP 213  
 Db 236 qpeppspmaeeelaefpeptwpbgelgphaglaglteesedvdallaeaavgrkrkwssp 295  
 Qy 214 LPPSWHIKQLYLRHVDQMSFGMTPLPAFAGHV-----PEATRVFPQVNVTKMG 266  
 Db 296 srsifhfg---rhlpdepaelglervkasvenhrlkgrpegiekgpp---rrkp 349  
 Qy 267 SWGHFNCSYCSFLAPEDPIFFIIGSLFLRELKIFERTDHIYGADTFNEMQPPSSPSY 326

Db 350 glasfirlsglkswdrap-----tfirel-----sdevv 378  
 Qy 327 LAAATTAATVYEAMTAVDTEAVWLLOGLWLOHQQFQWGPQAQIRAVLCAVPRG--RLVLVDLF 384  
 Db 379 lqgsvtlacqvsaqpaadatwskdgaple-----ssrvlisatlkfnqllvllvv 430  
 Qy 385 AESQPVYTRTASFQGPFIWCLMHNFGNGHGLFGALEAVNGGPEARLFPNSTMVGTGMA 444  
 Db 431 aedlgvyt-----csvsnalgtvtttgvlr-----kaer--psss----- 463  
 Qy 445 PEGISONEVWYSLMAELGMRKDPVPD-----LAAW--VTSPAARYGVSHPDAGAA 493  
 Db 464 -----pcpdigevyadgvlvkwkpvesygpytyivqcsleggs 501  
 Qy 494 WRLLLRSVYNC-----SGEAC 509  
 Db 502 wttlasdifdcyltsklrsgtytfrtactvskagmgypssseqvllggpshlasees 561  
 Qy 510 RGHNRSPVLRPSLOMNTSIWYNR-SDYFEAWRLLLTSAPSLATSPAPRYDLDLTR--Q 566  
 Db 562 qgrsaqplpstktkfafqtqgrgrfsvrwcwkasgralaakiipypkdktaavlreye 621  
 Qy 567 AVQELVSLYBEARSAYLS-KELASLLRAGGVLAAYELLPALDEVLASDSRELLGSWLEQA 625  
 Db 622 alkgirhphlaqlhaaylsprhlvllel---csgpellplclae-rasyseesevkdylwqm 678  
 Qy 626 RAAA-----VSEAEADFYEQN---SRVOLTLMGPENILDYANKLAGLVANYTTPRWRL 677  
 Db 679 lsaqtqylhnqhlhlrlrseumilteynll-----kvvdlnaqslsqekvlpsdkfk 732  
 Qy 678 FLEALVDSVAOG---IPQOQHQDKNVFOL-EQAFVLSKQRYPPQPRG 721  
 Db 733 yletnapellegqavp-----qtdlwaigvtafimlisaeypvsseg 774  
 RESULT 8  
 AAB30567  
 ID AAB30567 standard; Protein; 1351 AA.  
 XX  
 AC AAB30567;  
 XX  
 DT 19-MAR-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human signal transduction polypeptide.  
 XX  
 KW Signal transduction: H19G5; kinase; cardiac disease; angina pectoris;  
 KW congestive heart failure; dilated congestive cardiomyopathy;  
 KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
 KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
 KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
 KW atherosclerosis; cardiac tumour; microbial infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200063381-A1.  
 XX  
 PD 26-OCT-2000.  
 XX  
 PF 11-APR-2000; 2000WO-US09488.  
 XX  
 PR 16-APR-1999; 99US-0129553.  
 XX  
 XX (SCIO-) SCIOS INC.  
 PA  
 XX Zeng W, Stanton L, Kong H;  
 PI  
 XX WPI; 2001-007013/01.  
 DR  
 DR N-PSDB; AAC62285.  
 XX  
 PT Novel h19G5 polypeptides capable of regulating signal transduction and  
 PT exhibiting kinase activity useful for identifying antibodies to treat  
 PT cardiac diseases, and additional mediators of signal transduction -  
 XX

Claim 1; Page 55-57; 81pp; English.

The present sequence represents a human protein with putative function in signal transduction. The polypeptide is designated H19G5. The protein is capable of regulating signal transduction and exhibits kinase activity. The H19G5 transcript is expressed in the heart. H19G5 polypeptides and polynucleotides are useful for preventing or treating a cardiac disease, such as congestive heart failure, dilated congestive cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve disease, aortic valve disease or tricuspid valve disease, angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary arterial or renovascular hypertension, arteriosclerosis, atherosclerosis and cardiac tumours in humans. The polypeptide is also useful for detecting the expression of a protein capable of regulating signal transduction or the expression of a protein capable of acting as a donor or acceptor molecule of a phosphate group. The monoclonal antibodies can be used as probes for detecting discrete antigens expressed by tissue or cell samples, and therefore used in humans for localization and monitoring of microbial infection.

Sequence 1351 AA;

erv Match 3.08; Score 118.5; DB 22; Length 1351;

1st Local Similarity 18.6%; Pred. No. 0.48;

tches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

19 GGAAGDEAREAAVRLVARL-LGPGPAAD--FSVSVERALAAKPGLD-----TY 65

501 qasqasssqvsslrvgssqvqtepgpsldaegwtqeaedlsdstptlqrpqeqatmrkf 560

66 SLGG-GGAARVRVGSTGVAAAGLHRYLRDFCG-----CHVAWSGSQLR----- 109

561 slqrrqqvavgygtfagggdag-----gmlgggpmwariawvsqseeeqeeareae 614

110 "-----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSYSFVWWDWARWE 151

615 sqseeqqearaesplpqvsarpvgravptrssp-----eptpwe 655

152 -----REI--DWMALNGINLAL-----AWSQEAIWQRVYLAL----- 182

656 digqvsivqirdlsgdaeaadtisldisevdpaylnlsdlydkylpfefmifrkvpsa 715

183 -----GLTQAEINEF-----FTGP-----AFLAWGRMGNLHTWDGP 213

716 qppeppmaeeelaefpeptwpwpgelgphaggleiteesedvdallaeaavgrkrkwssp 775

214 LPPSWHIKQLYLQHRVLDQMRSEGMTPLPAFAGHV-----PEAVTRVFQPQVNVTKMG 266

776 srslfhpg---rhlpldepaelglrervkasvehisrilkgprpeglekegpp--rkkp 829

267 SWGHFNCSYSCSFLAPEDPIFPIIGSLFLRELKEFGTDHIYGADTFNEMQPPSSEPSY 326

```
830 qlasfrlsqkswdrap-----tflrel-----sdetv 858
```

327 LAAATTAVYEAMTAVDTEAVWLLOGWLFQHQPOFWGPAQIRAVLGAVPRG--RLVLDLF 384

859 laasvtlacqvsaaqatwskdqaple-----sssrvlisatlknfalltilvvv 910

385 AESOPVYTRTASFQOGOPFIWCMLHNFEGGNHGLFGALEAVNGGPEAARLFPNSTMVGTMGA 444

911 aedlavvt-----csvsnalqtvtttqvlr-----kaer--psss----- 943

445 PEGTSONEVVYSLMAELGWRKDPVPD-----LAAW--VTSFAARRYGVSHPDAGAA 493

944 -----pcpdigevvadqvlvkwpvesyqpvtviqcsleggs 981

494 WRITJRSVNC-----SGEAC 509

982 wttl asd l f d c c v l t s k l s r g a t v t f r t a c v s k a q m q p y s s p s e q v l l q q p s h l a s e e s 104

510 RGHNRSPVRRPSLOMNTSIWYNR-SDVFEAWRLLTSAPSLATSPAIFYDLDLTR--Q 566

[illegible]

|    |      |   |       |
|----|------|---|-------|
| Db | 1042 | qgrsaqlptsktffaftqtdiqgrfsvrrqcwckasralaakliip/hpdkdttavlreye | 11010 |
| Qy | 567  | AVQELVSLYEEARSAYLS-KELASLLRAGGVLAYELPALDEVLASDSRFLGSLWEQA     | 625   |
| Db | 1102 | alqrlphlaqlhaayisprhlvllel-csgpellpciae-rasyesevkdylwqm       | 1158  |
| Qy | 626  | RAAA-----VSEAEADFYEQN---SRYQLTLGPGENILLYANKQLAGLVANNYYTPRWL   | 677   |
| Db | 1159 | lsatqylhnhqhlhlldrsmnllteynll-----kvvdlnqagslsqekvlpssdkfkd   | 1212  |
| Qy | 678  | FLEALVDVNAQG---IPFOHQFDKNVQL-EQAEVLSKQRYPSQPRG                | 721   |
| Db | 1213 | vlermapelllegacvvp-----qtdlwaigvtatimlasaeypsvsseg            | 1254  |

## RESULT 9

|         |  |
|---------|--|
| AB30568 |  |
| ID      | AB30568 standard; Protein; 1610 AA.                                      |
| XX      |  |
| AC      | AB30568;   |
| XX      |  |
| DT      | 19-MAR-2001 (first entry)  |
| XX      |  |
| DE      | A full length human signal transduction polypeptide.                     |
| XX      |  |
| KW      | Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;    |
| KW      | congestive heart failure; dilated congestive cardiomyopathy;             |
| KW      | hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;   |
| KW      | mitral valve disease; aortic valve disease; tricuspid valve disease;     |
| KW      | myocardial infarction; cardiac arrhythmia; arteriosclerosis;             |
| KW      | atherosclerosis; cardiac tumour; microbial infection.                    |
| XX      |  |
| OS      | Homo sapiens.  |
| XX      |  |
| PN      | WO200063381-A1.  |
| XX      |  |
| PD      | 26-OCT-2000.   |
| XX      |  |
| PF      | 11-APR-2000; 2000WO-US09488.   |
| XX      |  |
| PR      | 16-APR-1999; 99US-0129553.   |
| XX      |  |
| PA      | (SCIO-) SCIOS INC.   |
| XX      |  |
| PI      | Zeng W, Stanton L, Kong H;   |
| XX      |  |
| WPI     | 2001-007013/01.  |
| DR      | N-PSDB; AAC62286.  |
| DR      |  |
| XX      |  |
| PT      | Novel h19G5 polypeptides capable of regulating signal transduction and   |
| PT      | exhibiting kinase activity useful for identifying antibodies to treat    |
| PT      | cardiac diseases, and additional mediators of signal transduction -      |
| XX      |  |
| PS      | Claim 1; Page 61-65; 8lpp; English.                                      |
| XX      |  |
| CC      | The present sequence represents a human protein with putative function   |
| CC      | in signal transduction. The polypeptide is designated H19G5. The protein |
| CC      | is capable of regulating signal transduction and exhibits kinase         |
| CC      | activity. The H19G5 transcript is expressed in the heart. H19G5          |
| CC      | polypeptides and polynucleotides are useful for preventing or treating a |
| CC      | cardiac disease, such as congestive heart failure, dilated congestive    |
| CC      | cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, |
| CC      | mitral valve disease, aortic valve disease or tricuspid valve disease,   |
| CC      | angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,   |
| CC      | arterial or rheovascular hypertension, arteriosclerosis, atherosclerosis |
| CC      | and cardiac tumours in humans. The polypeptide is also useful for        |
| CC      | detecting the expression of a protein capable of regulating signal       |
| CC      | transduction or the expression of a protein capable of acting as a donor |
| CC      | or acceptor molecule of a phosphate group. The monoclonal antibodies can |
| CC      | be used as probes for detecting discrete antigens expressed by tissue or |
| CC      | cell samples, and therefore used in humans for localization and          |
| CC      | monitoring of microbial infection.                                       |
| XX      |  |

|    |          |   |
|----|----------|---|
| SQ | Sequence | 1610 AA;  |
| XX | DT       | 25-SEP-2001 (first entry)   |
| XX | DE       | Human protein kinase SGK145.  |
| XX | KW       | Protein kinase; enzyme; cytostatic; nootropic; neuroprotective; human; antiparkinsonian; virucide; antibacterial; antifungal; antimigraine;   |
| XX | KW       | analgic; hypotensive; hypertensive; immunosuppressive; antiallergic;  |
| XX | KW       | antipsoriatic; antirheumatic; antiarthritic; ophthalmologic; anorectic;   |
| XX | KW       | osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;   |
| XX | KW       | vasotropic; antidiabetic; gene therapy.   |
| OS | OS       | Homo sapiens.   |
| XX | PN       | WO200155356-A2.   |
| XX | PD       | 02-AUG-2001.  |
| XX | PF       | 25-JAN-2001; 2001WO-US02337.  |
| XX | PR       | 25-JAN-2000; 2000US-0178078.  |
| XX | PR       | 31-JAN-2000; 2000US-0179364.  |
| XX | PR       | 17-FEB-2000; 2000US-0183173.  |
| XX | PR       | 17-MAR-2000; 2000US-0190162.  |
| XX | PR       | 29-MAR-2000; 2000US-0193404.  |
| XX | PR       | 13-NOV-2000; 2000US-0247013.  |
| XX | PA       | (SUGE-) SUGEN INC.  |
| XX | PI       | Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;  |
| XX | DR       | WPI; 2001-476202/51.  |
| XX | DR       | N-PSDB; AAH46904.   |
| XX | XX       | Kinase polypeptides useful for treating cancers, Alzheimer's disease, viral infections, diabetes, obesity, organ transplant rejection and rheumatoid arthritis -  |
| XX | PS       | Claim 7; Page 215; 218pp; English.  |
| XX | CC       | The invention provides human protein kinases and protein kinase-like enzymes and polynucleotides encoding the polypeptides. The kinase polypeptides and their modulators are useful for treating a disease or disorder such as cancer, immune-related diseases, cardiovascular disease, brain or neuronal-associated disease and metabolic disorders, including cancers of tissues, cancers of hematopoietic origin, diseases of the central nervous system, diseases of the peripheral nervous system, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis, viral infections, infections caused by prions, bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction, mood disorders, attention disorders, cognitive disorders, hypotension, hypertension, psychotic disorders, neurological disorders, dyskinesias, metabolic disorders, and organ transplant rejection. They are also useful for treating rhinitis, asthma, chronic inflammatory pelvic disease, chronic osteoarthritis, asthma, chronic inflammatory arthritis, metabolic disorders such as diabetes, obesity, cardiovascular diseases such as reperfusion injury, coronary thrombosis, clotting disorders and atherosclerosis, ocular diseases such as glaucoma, retinopathy and macular degeneration, psychiatric and neurological disorders such as anxiety, schizophrenia, dementia, manic depression, etc. The polynucleotides are useful in gene therapy techniques to treat the above mentioned disorders. Sequences AAB85491-85522 represent the human protein kinases of the invention. |
| XX | XX       | Sequence 1618 AA;   |
| XX | XX       | Query Match 3.0%; Score 118.5; DB 22; Length 1618;  |
| XX | XX       | Best Local Similarity 18.6%; Pred. No. 0.63;  |
| XX | XX       | Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;  |
| QY | 19       | GGAGDEAREAAVRAVLVRL-LGPGPAAD---FVSVERALAAKPGCLD-----TY 65   |
| Db | 760      | gsasqassvslrgvssqvgteppslsdaegwtgaedlsdtpclrpqgeatmrkf 819  |
| QY | 66       | SLGG-GGAARVRVGTGVAAGLHRYLRDFCG----CHVAMSGSOLR----- 109  |
| Db | 820      | slgrggyagvdygtfagdgag-----gmllggpmwariawavsqseeeqearae 873  |
| QY | 110      | -----LP-----RPLPAVPGELTEATPNRYRYONVTCOSYFVWWDWARWE 151  |
| Db | 874      | sqseeqearaesplqvsaarpvpgvgraptrssp-----eptpwe 914   |
| QY | 152      | -----REI--DWMALNGINAL-----AMSGOEAIWRVYLAL----- 182  |
| Db | 915      | dggvslvqlrdlsgdaeaadtisldisevdpaylnlsdlydkylpftfmfrkvpksa 974   |
| QY | 183      | -----GLTQAEINER-----FTGP-----AFLAWGMGNLHTWDGP 213   |
| Db | 975      | qppeppmaeeelaefpeptwpggelgphagleiteesedvdallaeaavgrkrkwssp 1034   |
| QY | 214      | LPPSWHIKOLYLOHRYLDQMRSEGMTPLPAFAGHV-----PEAVTRVFQVNVTKMG 266  |
| Db | 1035     | srsalfhfpq---rhpldepaeglrvkvasvehisrllkgrpeglekpgp---rrkbp 1088   |
| QY | 267      | SWGHFNCSYSCSCLLAPEDIPFIIGSLFLRELIIKEFGTDHYGADTFNEMQPPSEPSY 326  |
| Db | 1089     | glasrlsglkswdrap-----tflrel-----sdetvv 1117   |
| QY | 327      | LAAATTAATVAMTAVTEAWVLQGLWLFQHPQFWGPAQIRAVLGAIVPRG---RLVLDLF 384   |
| Db | 1118     | lqgsvtlacqvsqaqaatgskdgaple-----ssrvlisatllknflltilvv 1169  |
| QY | 385      | ASQPVYTTASFGQGFITWCLHNFNGHGLFGALEAVNGGPEAARLPNSTVMVGTGMA 444  |
| Db | 1170     | aedlgyvt-----csvsnalgtvtgtvlr-----kaer--psss----- 1202  |
| QY | 445      | PEGISQNEVVYSILMAELGWRKQDPVD-----LAAW--VTSPAARYGVSHPDAGAA 493  |
| Db | 1203     | -----pcpdigevyadgvlvkvkpesygpvtvlyvcslleggs 1240  |
| QY | 494      | WRLLSRVYNC-----SGEAC 509  |
| Db | 1241     | wtllasdifdcyltsklrgtgyfrtaacvskagmgypsspgvllgapshlasees 1300  |
| QY | 510      | RGNRSPLVRRPSLOMNTSIWNR--SDVFEAWRLLLTSAPSLSATSPAFRYDILLDLTR--Q 566   |
| Db | 1301     | qgrsaqplpstktfatqtlqgrfsvvrqwcakasgralaakliipypkdktaavlreye 1360  |
| QY | 567      | AVQELVSLYEARSAYLS-KELASILLRAGGVLAPELLPALDEVLASDRFLGSLWLSQA 625  |
| Db | 1361     | alkgrhphlaqhaaylsprhlvllel--csgpellipclae-rasysevekdylwqm 1417  |
| QY | 626      | RAAA-----VSPAEDAFYEON---SRYQITLWGPPEGNILDVANKLAGLVANYTTPRWRL 677  |
| Db | 1418     | lsatylhnhqhlidlrsemmiiteynll-----kvvdignadslsqekvlpsdkfk 1471   |
| QY | 678      | FLEALVDSVAQG---IPFQHQHFDKNVQL-EQAFVLSKQRYPSQPRG 721   |
| Db | 1472     | yietmapelledggavp-----qtdiwaigvtafimlsaeypvsseg 1513  |

RESULT 10

AAB85504

ID AAB85504 standard; protein; 1618 AA.

XX

AC AAB85504;

|          |             |  |      |    |  |
|----------|-------------|--|------|----|--|
| Db       | 768         | gasqasseqvaslrvsgsqvgtpeppsldeagwtqeaedlsdstptlqrpqeqatmrkf  | 827  | KW | myocardial infarction; cardiac arrhythmia; arteriosclerosis; atherosclerosis; cardiac tumour; microbial infection; splice variant. |
| QY       | 66          | SLGG-GGAARVRVSGSTGVAAAGLHRYLRDFCG-----CHVAMSGSQLR-----   | 109  | XX | Homo sapiens.  |
| Db       | 828         | slggrggyagvagytgafaggdag-----gmlgggpmwariawavsqseeeqearae  | 881  | OS | Key  |
| QY       | 110         | -----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSYFVWMDWARWE   | 151  | FH | Domain   |
| Db       | 882         | sqseeqearaesplpqvsarpvpevgraptrssp-----eptpwe  | 922  | FT | Location/Qualifiers  |
| QY       | 152         | -----REI-----DWMALNGINLAL-----AWSQCEAIWQRYLYAL-----  | 182  | FT | 325..504   |
| Db       | 923         | diggvslvirdlsgdaeaadtisidsevdpaylnlsdlydikylptefmifkrkvpksa  | 982  | FT | /note= "guanine nucleotide exchange factor domain"   |
| QY       | 183         | -----GLTQAEINER-----FTGP-----AFLANGRMGNLHTWDGP   | 213  | FT | 1094..1351   |
| Db       | 983         | qpeppspmaeeelaefpeptwpgelgphaglaiteesedvdallaavaavgrkrkwasp  | 1042 | FT | /note= "kinase domain"   |
| QY       | 214         | LPPSWHIKQLYQHRYLDQMRSEGMTPLPAFAGHV-----PEAVTRVFQVNVTKMG  | 266  | FT | 2301..2553   |
| Db       | 1043        | srslfhfgp---rhlpdpeaelgrrerkasvehisrlkgrpeglekgpp---rkkp   | 1096 | FT | /note= "kinase domain"   |
| QY       | 267         | SMGHFNCSYCSFLAPEDPIPIIGSLFLRELKEFGTHIYGADTFNEMQPPSPSEPSY   | 326  | FT |  |
| Db       | 1097        | glasfrlsglkswdrap-----tflrel-----sdtv  | 1125 | FT |  |
| QY       | 327         | LAATTAAYEAMTAVTEAVWLLOGLWFQHPQFQWGPQAIQRAVLGAVPRG--RLVL  | 384  | FT |  |
| Db       | 1126        | lgsvtlacqvsqaqaatwskdgaple-----ssrvliisatlknflltlivvv  | 1177 | FT |  |
| QY       | 385         | AESQPVVTRTASQGOPTWCMILHNEGGNHLFGALEAVNGPEAARLPNSTMTVGTGMA  | 444  | FT |  |
| Db       | 1178        | aedlgvyt-----cvsvalgtvtttgvlr-----kaer--psas-----  | 1210 | FT |  |
| QY       | 445         | PEGISQNEVVYSLMAELGWRKDPDP-----LAAW--VTSFARRYGVSHPDAGAA   | 493  | FT |  |
| Db       | 1211        | -----pcpdigevyadgvlvwkpvesygpvtiyvqcsleggs   | 1248 | FT |  |
| QY       | 494         | WRLLLRSVYNC-----   | 509  | FT |  |
| Db       | 1249        | wttlaedifccyltsklrggtytfrtacyskagmpyspseqvllggpshlaseees   | 1308 | FT |  |
| QY       | 510         | RGNRSPLVRRPSLQMTWSIWNR-SDVFEAWRLLTSAFRLTSAFRLDLDLFR--Q   | 566  | FT |  |
| Db       | 1309        | qgrsaqlpstktfatqtdqrfsrvrcweskagralaakliphpkdkttavireye  | 1368 | FT |  |
| QY       | 567         | AVQELVSLYYEARSAYLS-KELASLLRAGGVLAYELLPALDEVLASDSRFLLCGLEQA   | 625  | FT |  |
| Db       | 1369        | aikglrphlaqlhaayisphvllel--csgpellpciae-rasysevekdylwqm  | 1425 | FT |  |
| QY       | 626         | RAAA-----VSEAEADFEQN---SRYQLTLWGPENGLDANKQLAGLVANYTTPRWRL  | 677  | FT |  |
| Db       | 1426        | lsatqylhnhqhlhldirsenmitleynll-----kvvldlgnagslsqekvlpsdkfk  | 1479 | FT |  |
| QY       | 678         | FLEALVDSVAGQ---IPFQHQFDKNVQL-EQAFVLSKORYPSQPRG   | 721  | FT |  |
| Db       | 1480        | yletmepellleggavp-----qtdiwaigvtafimlmsaeypvsseg   | 1521 | FT |  |
| RESULT   | 11          |  |      |    |  |
| AAB30569 |             |  |      |    |  |
| ID       | AAB30569    | standard; Protein: 2596 AA.  |      |    |  |
| XX       | AAB30569;   |  |      |    |  |
| AC       |             |  |      |    |  |
| XX       |             |  |      |    |  |
| DT       | 19-MAR-2001 | (first entry)  |      |    |  |
| XX       |             |  |      |    |  |
| DE       | A           | splice variant of a signal transduction polypeptide.   |      |    |  |
| XX       |             |  |      |    |  |
| KW       |             | Signal transduction; H19G5; kinase; cardiac disease; angina pectoris; congestive heart failure; dilated congestive cardiomyopathy; hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension; mitral valve disease; aortic valve disease; tricuspid valve disease; |      |    |  |



351 ahmpttavihevqrf-----gdivpl 372

RESULT 13

AAR72377

ID AAR72377 standard; Protein; 497 AA.

XX

AC AAR72377;

XX

DT 15-NOV-1995 (first entry)

XX

DE Human auxillary cytochrome P450 species 2D6 variant 2 protein.

XX

KW Human cytochrome P450; amplification; PCR; primer; expression vector;

XX

KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;

XX

KW carcinogen; mutagen; liver metabolism.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT Misc-difference 296 /note= "Cys to Arg variation"

FT

FT Misc-difference 486 /note= "Thr to Ser variation"

XX

XX EP644267-A.

XX

PN 22-MAR-1995.

PD

PD 20-JUL-1994; 94EP-0111298.

XX

XX 21-JUL-1993; 93JP-0180246.

XX

PR 20-JUL-1993; 93JP-0201120.

PR

PR 30-JUL-1993; 93JP-0208279.

XX

XX (HAYASHI) HAYASHI K.

PA (SUMO ) SUMITOMO CHEM CO LTD.

XX

XX Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;

XX

PI Yabusaki Y;

PI

PI WPI: 1995-116991/16.

DR N-PSDB; AAQ87731.

XX

XX Evaluation of safety of a chemical cpd. - using recombinant yeast

PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase

PT

XX

XX Examples; Page 91-93; 124pp; English.

XX

XX The amino acid sequence of the human auxillary cytochrome P450 species

CC 2D6 variant 2. This variant contains variations at residues 296: Cys to

CC Arg and 486: Thr to Ser, caused by variations at bases 886: T to C and

CC 1457: C to G in the DNA sequence. The cDNA was amplified by PCR using

CC the primers AAQ87763-6. The product was cloned into the yeast

CC expression vectors pAAH5N or pAHRP to produce the vectors p2D6 variant 2

CC for the expression of the cytochrome P450 alone or p2D6R variant 2 for

CC co-expression with the yeast NADPH-P450 reductase.

CC The vectors are used in a method for evaluating the safety of a chemical

CC compound by reacting the chemical compound with recombinantly produced

CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),

CC 2E1 (AAQ87716), or 3A4 (AAQ87718), or their auxiliary species and

CC variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused

CC protein or in cell extracts, and analysing the resulting metabolite to

CC assess the safety of the chemical compound. The method is useful for

CC determining whether the chemical compound, or its metabolite, will be

CC converted into a carcinogenic or mutagenic form through metabolism in the

CC liver.

XX

XX Sequence 497 AA;

XX

2



Db 351 ahmpyttavihevqrf-----gdivpl 372

Search completed: September 21, 2002, 12:26:27  
Job time: 162 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2002, 12:23:45 : Search time 14.54 Seconds  
(without alignments)  
1248.158 Million cell updates/sec

Title: US-09-836-613-2  
Perfect score: 3939  
Sequence: 1 MEAVAVAAAVGVLLAGAGG.....VDLAKKIFLKYPGWAGSW 743

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description       |
|------------|-------|---------------|--------|-------|-------------------|
| 1          | 3939  | 100.0         | 743    | 4     | US-09-077-354B-2  |
| 2          | 121   | 3.1           | 4472   | 2     | US-08-804-227C-2  |
| 3          | 120   | 3.0           | 3729   | 2     | US-08-804-227C-4  |
| 4          | 113   | 2.9           | 7257   | 3     | US-09-335-409-5   |
| 5          | 113   | 2.9           | 7257   | 4     | US-09-568-102-5   |
| 6          | 113   | 2.9           | 7257   | 4     | US-09-567-969-5   |
| 7          | 113   | 2.9           | 7257   | 4     | US-09-568-480-5   |
| 8          | 113   | 2.9           | 7257   | 4     | US-09-568-486-5   |
| 9          | 113   | 2.9           | 7257   | 4     | US-09-568-472-5   |
| 10         | 107.5 | 2.7           | 1621   | 4     | US-08-972-927-3   |
| 11         | 107.5 | 2.7           | 1622   | 4     | US-08-972-927-6   |
| 12         | 105   | 2.7           | 1203   | 4     | US-09-075-272-4   |
| 13         | 105   | 2.7           | 4150   | 4     | US-09-428-517-2   |
| 14         | 104.5 | 2.7           | 726    | 2     | US-08-313-185-49  |
| 15         | 104.5 | 2.7           | 726    | 2     | US-08-459-499-13  |
| 16         | 104.5 | 2.7           | 726    | 3     | US-09-082-614A-49 |
| 17         | 104   | 2.6           | 581    | 3     | US-09-221-235-8   |
| 18         | 104   | 2.6           | 581    | 3     | US-09-221-928-8   |
| 19         | 104   | 2.6           | 581    | 4     | US-09-221-527-8   |
| 20         | 104   | 2.6           | 581    | 4     | US-09-221-236-8   |
| 21         | 104   | 2.6           | 581    | 4     | US-09-221-416-8   |
| 22         | 104   | 2.6           | 581    | 4     | US-09-221-245-8   |
| 23         | 104   | 2.6           | 581    | 4     | US-09-163-115-8   |
| 24         | 104   | 2.6           | 581    | 4     | US-09-221-528-8   |
| 25         | 104   | 2.6           | 581    | 4     | US-09-593-553-8   |
| 26         | 104   | 2.6           | 581    | 4     | US-09-221-237-8   |
| 27         | 104   | 2.6           | 677    | 4     | US-08-836-567-2   |

|    |       |     |      |   |                   |
|----|-------|-----|------|---|-------------------|
| 28 | 104   | 2.6 | 1197 | 4 | US-08-836-567-12  |
| 29 | 101   | 2.6 | 3519 | 4 | US-09-428-517-4   |
| 30 | 101   | 2.6 | 5215 | 4 | US-09-105-537-2   |
| 31 | 100.5 | 2.6 | 2628 | 4 | US-09-413-814-11  |
| 32 | 100   | 2.5 | 1323 | 1 | US-08-026-138E-4  |
| 33 | 100   | 2.5 | 1841 | 2 | US-08-804-227C-6  |
| 34 | 99.5  | 2.5 | 719  | 1 | US-07-943-843-4   |
| 35 | 99.5  | 2.5 | 719  | 1 | US-08-347-003-4   |
| 36 | 99    | 2.5 | 729  | 2 | US-08-313-185-50  |
| 37 | 99    | 2.5 | 729  | 2 | US-08-459-499-14  |
| 38 | 99    | 2.5 | 729  | 3 | US-09-082-614A-50 |
| 39 | 99    | 2.5 | 1230 | 2 | US-08-968-542C-35 |
| 40 | 98.5  | 2.5 | 670  | 2 | US-08-473-750-11  |
| 41 | 98.5  | 2.5 | 670  | 2 | US-08-477-326-11  |
| 42 | 98    | 2.5 | 1611 | 2 | US-08-804-227C-5  |
| 43 | 97    | 2.5 | 1336 | 2 | US-08-231-193A-58 |
| 44 | 97    | 2.5 | 1336 | 2 | US-08-486-273A-58 |
| 45 | 97    | 2.5 | 1336 | 3 | US-08-940-086A-58 |

ALIGNMENTS

RESULT 1  
US-09-077-354B-2  
; Sequence 2, Application US/09077354B  
; Patent No. 6255096  
; GENERAL INFORMATION:  
; APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
; APPLICANT: WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART  
; TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
; TITLE OF INVENTION: -N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SA  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
; STREET: 400 GARDEN CITY PLAZA  
; CITY: GARDEN CITY  
; STATE: NEW YORK  
; COUNTRY: UNITED STATES  
; ZIP: 11530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/09/077,354B  
; APPLICATION NUMBER: US/09/077,354B  
; FILING DATE: 22-APRIL-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/00747  
; FILING DATE: 22-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: POKALSKY, ANN R.  
; REGISTRATION NUMBER: 34,697  
; REFERENCE/DOCKET NUMBER: 12416  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 516 742 4343  
; TELEFAX: 516 742 4366  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 743 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Potentially-glycosylated Asn site,  
; LOCATION: 261  
; FEATURE:  
; NAME/KEY: Potentially-glycosylated Asn site,  
; LOCATION: 272  
; FEATURE:  
; NAME/KEY: Potentially-glycosylated Asn site,  
; LOCATION: 435

FEATURE: Potentially-glycosylated Asn site,  
NAME/KEY: 503  
LOCATION:  
FEATURE: Potentially-glycosylated Asn site,  
NAME/KEY: 513  
LOCATION:  
FEATURE: Potentially-glycosylated Asn site,  
NAME/KEY: 526  
LOCATION:  
FEATURE: Potentially-glycosylated Asn site,  
NAME/KEY: 532  
LOCATION:  
US-09-077-354B-2

Query Match 100.0%; Score 3939; DB 4; Length 743;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAVAVAAAGVLLLAGAGGAAGDEAREAAAVRALVRLGPGPAADFSVVERALAAKP 60  
DB 1 MEAVAVAAAGVLLLAGAGGAAGDEAREAAAVRALVRLGPGPAADFSVVERALAAKP 60  
QY 61 GLDYSLGGGGAARVRVRSSTGVAAAGLHRYLRDFCGCHVAMSGSQLRPLRPVAPGE 120  
DB 61 GLDYSLGGGGAARVRVRSSTGVAAAGLHRYLRDFCGCHVAMSGSQLRPLRPVAPGE 120  
QY 121 LPEATPNRYRYONVCTOSYSFVWMDWARWEREIDNMALNGINLAWSGOEAIWQVYL 180  
DB 121 LPEATPNRYRYONVCTOSYSFVWMDWARWEREIDNMALNGINLAWSGOEAIWQVYL 180  
QY 181 ALGTOAEINEFTGPAFLAWGMGNLHTWDGPLPSPWHIKOLYLQHRVLDQMRSGMTP 240  
DB 181 ALGTOAEINEFTGPAFLAWGMGNLHTWDGPLPSPWHIKOLYLQHRVLDQMRSGMTP 240  
QY 241 VLPFAAGHVPEAVTRVPPQNVYTKMSGWHFNCYSYCSFLLAPEDPIFFIIGSLFREL 300  
DB 241 VLPFAAGHVPEAVTRVPPQNVYTKMSGWHFNCYSYCSFLLAPEDPIFFIIGSLFREL 300  
QY 301 KEFTGTHYAGDTFNMOPSPSSPSYLAATAATVAYEAMTAVDTEAVNLLQGWLFQHPQF 360  
DB 301 KEFTGTHYAGDTFNMOPSPSSPSYLAATAATVAYEAMTAVDTEAVNLLQGWLFQHPQF 360  
QY 361 WGPQAIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFQGPFFIWCMLHNFNGHGLFGAL 420  
DB 361 WGPQAIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFQGPFFIWCMLHNFNGHGLFGAL 420  
QY 421 EAVNGGPEAARLPNSTWVGTMGMAPEGISQNEVVYSLMAELGWRKDPVPLAAWYTSFAA 480  
DB 421 EAVNGGPEAARLPNSTWVGTMGMAPEGISQNEVVYSLMAELGWRKDPVPLAAWYTSFAA 480  
QY 481 RRYGVSHPDAGAAWRLLLRSVYNGSGEACRGNRSPVLRPRSLQMNSTSIWYNRSDVFEAW 540  
DB 481 RRYGVSHPDAGAAWRLLLRSVYNGSGEACRGNRSPVLRPRSLQMNSTSIWYNRSDVFEAW 540  
QY 541 RLLTSAPSLATSPAFRYDLDLTRQAVQELVSLYFEARSAYLSKELASLLRAGGVLAY 600  
DB 541 RLLTSAPSLATSPAFRYDLDLTRQAVQELVSLYFEARSAYLSKELASLLRAGGVLAY 600  
QY 601 ELLPALDEVLASDSRFLGSLWLEQARAAVSEAEADFYEQNSRYQLTLWGPENILYAN 660  
DB 601 ELLPALDEVLASDSRFLGSLWLEQARAAVSEAEADFYEQNSRYQLTLWGPENILYAN 660  
QY 661 KQAGLVANYTTPRWLFLEALVDSVAGQIPFQHQFDKNVQLEQAFVLSKQRYPSQPR 720  
DB 661 KQAGLVANYTTPRWLFLEALVDSVAGQIPFQHQFDKNVQLEQAFVLSKQRYPSQPR 720  
QY 721 GDTVDLAKKIIFLYPGWAGSW 743  
DB 721 GDTVDLAKKIIFLYPGWAGSW 743

RESULT 2

US-08-804-227C-2  
; Sequence 2, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804.227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4472 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-804-227C-2

Query Match 3.1%; Score 121; DB 2; Length 4472;  
Best Local Similarity 21.6%; Pred. No. 0.19;  
Matches 176; Conservative 71; Mismatches 280; Indels 288; Gaps 44;

QY 4 VAVAAV-----GVLLLAGAGGAAGDEAREAAVRAVRLG-----PGPAAD 47  
DB 1686 VDAVAVVGHSGSIAAATVAGALSLE--DAAAVVALRAGLIGRYLAGRGMAAVPLPAGE 1743  
QY 48 FSVSVERALAAKPGLDYSLGGGAA-----RVRVRGSTGVAAAAGLH----- 90  
DB 1744 ----VEAGLAKWPGVEVAAGVAVGASTVVSQDRRAVAGYVAVCOAGVQARLIPVDYASHS 1799  
QY 91 RYLRFDCG-CHVAMSGSQLRPR-PLPAV-----PQELTEATPNRYRYONVCTOSYSFV 143  
DB 1800 RHVEDLKGELERVLGIRPRSPRPVPCSTVAGEQGPGEV----- 1838  
QY 144 WWDWARHER-----EIDWMLANGINL-----ALAWSGOEAIWQVYLA 181  
DB 1839 -FDAGYWFNRNLNRNRFSAVVVGGLEBEGHRRFTEVSAHPVLVHAIQTAEEA--DRSVHA 1895  
QY 182 LGLTQAE--INEFFTGPAPLAWGRMGNLHTWDCPLPPSWHIKOLYL-----QHRVLDOM 233  
DB 1896 TGTILRRDDSPHLLTSTA-EAWAHGATL-TWDPALPPG-HLTTLPYFNNHHYWLJ-- 1950  
QY 234 RSEGMTVPLPAFAGHVPEAVTRVPPQNVYTKMSGWHFNCYSYCSFLLAPEDPIFP-IIG 292  
DB 1951 ----TTPTTATTTSQPTDAWR-----YRVT-----WK-----ALTSSPVRPHSIG 1988  
QY 293 SLFLRELKKEFGTDHIYGADTFNEMOPSPSSPSYLAATAATVAYEAMTAV-----D 342  
DB 1989 RCLL-----VAPPTDGLLGLTTLVLSERGASVARLEVPICARR 2028

|    |      |  |                                  |      |
|----|------|--|----------------------------------|------|
| Qy | 343  | TEAVWLQGWLFQHQFQWGPQAQ----   | IRAVLGAVPRGRLLVLDFAESQPVYTRTASFO | 398  |
|    |      |  |                                  |      |
| Db | 2029 | AEVAELL-----KPSMESAGENTTVVSLGLVP-----                              | STDAVRTSIALQ                     | 2070 |
|    |      |  |                                  |      |
| Qy | 399  | -----GQPF--IWCMLE-----HFGGNGHGLFGALEAVN-----                       | GG-----PEAA                      | 430  |
|    |      |  |                                  |      |
| Db | 2071 | AVSDIGVPAARWALTRRAVAVVPGETPDQAGAQLWQGRVAALELPDIMGGLDLPETA          | 2130                             |      |
|    |      |  |                                  |      |
| Qy | 431  | RLFPNSTVMGTGMAPGEGISQ-----NEVVYSLMABELGWRKDPVPDLAAWTSTFAARBYGVSHYP | 488                              |      |
|    |      |  |                                  |      |
| Db | 2131 | ELTRTPETSPQPTPERLQTPNRRALELAAVYLAGRQDQOVAVRAGSIYGRVRSAAA           | 2190                             |      |
|    |      |  |                                  |      |
| Qy | 489  | DAGAAVR-----LLLRVYVNCSEACRGNRSPLVRRPSLQMTSIWYNRSDVFCAWRLLL         | 544                              |      |
|    |      |  |                                  |      |
| Db | 2191 | AGAAWQPSGTVLITGGMGAIGRR-----LARRLAAE-----GAERLVL                   | 2229                             |      |
|    |      |  |                                  |      |
| Qy | 545  | TS-----APSLATSPAFRYDLLD-----LTRQAVQELVSLY-----                     | 575                              |      |
|    |      |  |                                  |      |
| Db | 2230 | TSRRGPEAPCAA-----ELAEELRGHCGEVVHAACDVAERDAAALVATYPNVAVFT           | 2282                             |      |
|    |      |  |                                  |      |
| Qy | 576  | -----YEARSAYLSKELASLLRAGGVLAYELLPALDEVLASDRPLL-----                | GSWLEQARA                        | 627  |
|    |      |  |                                  |      |
| Db | 2283 | AGILDDAVDITLSPESFETVRGAKVCGAELLHQLTADIKGLDAFVLSPSSVTGTGNAGQG       | 2342                             |      |
|    |      |  |                                  |      |
| Qy | 628  | A-AVSEAEADFYEQNSR-----YQLTWLGPGE                                   | 653                              |      |
|    |      |  |                                  |      |
| Db | 2343 | AYANAALDALAERRRAAGLPATSVAMGLWGGGG                                  | 2377                             |      |

```

RESULT 3
US-08-804-227C-4
: Sequence 4, Application US/08804227C
: Patent No. 5876991
: GENERAL INFORMATION:
: APPLICANT: Dehoff, Bradley S.
: APPLICANT: Kustoss, Stuart A.
: APPLICANT: Rosteck, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) text only
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3729 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-804-227C-4

```

```
Query Match          3.0%; Score 120; DB 2; Length 3729;
Best Local Similarity 22.4%; Pred. NO. 0.18;
```

| Matches | 192: | Conservative   | 72:   | Mismatches                          | 265:                  | Indels                            | 328:               | Gaps |
|---------|------|--|---|-------------------------------------|-----------------------|-----------------------------------|--------------------|------|
| QY      | 4    | VAVAAAV  | ---   | GVLLLAGAGGAGGAEAREAAAVRALVARLLG     | ---                   | PGPAAD                            | 47                 |      |
| Db      | 668  | VDVAAVGVHSGEIAATAACVAGALSLE                            | ---   | DAAAVVALRAGLIGRYLAGRGMAAAVPLPAGE    | 725                   |                                   |                    |      |
| QY      | 48   | FSVSVERALAKPGLDITYSLGGGAA                              | ---   | RVRVGSTGVAAAAGLH                    | ---                   | 90                                |                    |      |
| Db      | 726  | ---  | VEAGLAKWPGVEYAAVNGPASTVVGSDRRAVAGYAVCAQEGVOARLIPVDYASHS | 781                                 |                       |                                   |                    |      |
| QY      | 91   | RYLRDFCG   | -CHVANGSGOLRUPR   | -PLPAV                              | ---                   | PGELTEATPNRYRYQNVCTQTSYFV         | 143                |      |
| Db      | 782  | RHVEDLKCELERVLSGIRPSR                                  | PRVPCVSTVAGEQGPV  | ---                                 | ---                   | 820                               |                    |      |
| QY      | 144  | WVDWARWR   | ---   | EIDWMALINGNL                        | ---                   | ALAWSGOEAIWQRYVLA                 | 181                |      |
| Db      | 821  | -FDAGYWFNRNLNRVEFS                                     | SAVVGGLLEBCHRRFEVSAHPVLVHAIEQTAEAA                      | -DRSVHA                             | 877                   |                                   |                    |      |
| QY      | 182  | LGLTQAE  | ---   | INEEFTGPAFLWGRGNLHTDGPPLPSPWHIKQLYL | ---                   | QHRVLDQM                          | 233                |      |
| Db      | 878  | TGTLRRQDDSPHRLLTSTA                                    | -EAWAHGATL  | -TWDPALPPG                          | -HUTTLTPTYFNFHHHWYLD  | ---                               | 932                |      |
| QY      | 234  | RSFGMTPLVPFAF  | ---   | AGHVPEAVT                           | -RVFPQVNVTKM          | -GSWGHHFNGSYCS                    | 278                |      |
| Db      | 933  | ---  | TTPTPATTTQSPDTAQNPADALPYKSVNKLRRQDOSTLARLDGRW           | ---                                 | ---                   | 978                               |                    |      |
| QY      | 279  | FLLAPEQPIPLIGSLFLRELK                                  | ---   | ---                                 | ---                   | EFGTDH                            | ---                | ---  |
| Db      | 979  | LLVYPEASADPSAEGVARELTARGATVESLTP                       | PEGADRSRLGLLVDATEDEAGPLRG                               | 103                                 |                       |                                   |                    |      |
| QY      | 323  | EPSYLA   | ---   | ---                                 | ---                   | AATTAVYEAMTAVDTEA                 | -VWLLQGLWLFQHPQFWG | 362  |
| Db      | 1039 | IVSLLLAGDHAGADGARPVVPAGLAASLALIQAGDAGTEAGLW            | ---   | ---                                 | ---                   | 108                               |                    |      |
| QY      | 363  | PAQTRAVILGAVPRGRLVLVDLFAESQPV                          | ---   | YTRTASFQGPQFICWMLHNFNGNHGLRG        | -418                  |                                   |                    |      |
| Db      | 1084 | -AVTRGAAVAPG   | ---   | DVPAPSOALLWGFGRVAGIE                | -LPCHWGGLLDLPTGPDGSGF | 1135                              |                    |      |
| QY      | 419  | ---  | ---   | ALEAVNGGPEAAARLFPNSTVMGT            | -GMAPGIGSONEVVYSLMAEL | 461                               |                    |      |
| Db      | 1136 | QLAATLAGRAPEDQVALRA                                    | -SGAYGRRLVRASAAGGADGWRPRG                               | ---                                 | ---                   | TVLVV                             | 1184               |      |
| QY      | 462  | GWRKDPVPLAANVTSAFARRY                                  | ---   | GVSHIP                              | ---                   | DAGAAWRLLLRSVYNCSEACRGHNR         | 514                |      |
| Db      | 1185 | GDTAEVAGVLVRWLLNGARRVTLTSLGSLP                         | PEELADVAARYTV   | ---                                 | ---                   | APCDPADR                          | 1236               |      |
| QY      | 515  | SPLVRRPSLQMNSTSIWNRSDFEAWRLLLT                         | SAPLSLATSPAFRYDLDDLTROA                                 | ---                                 | ---                   | 567                               |                    |      |
| Db      | 1237 | PALRTLLAEQAPTAV  | ---   | ---                                 | ---                   | LVAPPVPPTP                        | ---                | ---  |
| QY      | 568  | ---  | VOELSVLYE   | ---                                 | ---                   | FARSAYLSKELASLL                   | 592                |      |
| Db      | 1278 | KTGLVDRLLASLLDEPDPLLEDGEDLDAFVVFSSVAGVWGGAGGGVAACTAYLD | -ALAECR   | 1336                                |                       |                                   |                    |      |
| QY      | 593  | RAGG   | -VLAYELLPALDEVLASDRFLLSGLEQA                            | ---                                 | ---                   | RAAV                              | 630                |      |
| Db      | 1337 | RAGGLPVTSVANTPWLIG                                     | -TPAADS   | ---                                 | ---                   | LCEQMSRAGITPLDPAASLADALARAVRRAGCV | 1392               |      |
| QY      | 631  | SEAEADFFQNSRYQLT                                       | 647   |                                     |                       |                                   |                    |      |
| Db      | 1393 | TVADIDWERFASAYTAT                                      | 1409  |                                     |                       |                                   |                    |      |

RESULT 4  
US-09-335-409-5  
; Sequence 5, Application US/09335409  
; Patent No. 6131029  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon

## RESULT 4

REPORT #  
US-09-335-409-5

; Sequence 5, Application IIS/09335409

Patent No. 6121029

GENERAL INFORMATION:

APPLICANT: Schupp,

APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/335,409  
CURRENT FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-568-102-5

Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2.6;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSWHIKOLYLQHRVLDQMRSGMTVPVLPAPAGHVPEAVTRVFPQVNVTKM-GSWGHFN 272  
DB 4805 IPAEWPLQGVIIHAAGALDD-----GVLDEQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLLAPEDPIPIIGSLFLRELKFGTDHIYGADTFNEMQPPSPSEPSYLAATT 332  
DB 4847 ---NLHETAGNDLAFVLFSSMSGLLSGAGSQSNYAAANTFLDALAAHRAEGLAAQSL 4902

QY 333 A-----VYEAMTAVDTAEAVWLLQGWLFQHPQFQWGPQAIKRAVLG---AVPRGRL--L 379  
DB 4903 AWGPNWDGGMAGLSAA-----LQARLARHGMGALSAPQGTALLGQALARPETOLGAM 4955

QY 380 VLDLFAESQPVYTRTASFQGFPIWCMHLNFGNGHGLFGALEAVNGGPEAARLPNSTMV 439  
DB 4956 SLDVRAASQ-----ASGAAPVPMVRALVRAEARHTAAGAAGAL-----AARL----- 4997

QY 440 GTGMAPEGISONEVYSLMAE-----LQWR-----KDPVPDL-----AAWTSFAARR 482  
DB 4998 --GALPEARRADEVKRVQAEIARVLSWSAASAVPVDRLSDGLDLSLTAVELRNVLGR 5055

QY 483 YGVSS-----HPDAGAAWRLLLRSVYNCSEACRGNHRSPLV-----RRP 521  
DB 5056 VGATLPATLAFDHTVDALTRWLLDKVLAVAEPSVSSAKSSPQVALDEPIAIGICRFP 5115

QY 522 SLOMNTSIWNRSDVEAEWRLLTSAPSLATSPAFRYDLDLTLRQAVQLVSLYE---E 578  
DB 5116 G-----GVADPEFNRLLLEEGSDAVVEPHERWD-----IDAFYDPPDP 5154

QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSMLEQARAA 628  
DB 5155 VRGKMTI-----RFGGFLSDIDRDPAPFGISPREATTMDPQORLLLETSWEAPEFAG 5207

QY 629 AVSE 632  
DB 5208 ILPE 5211

RESULT 6  
US-09-567-969-5  
Sequence 5, Application US/09567969  
Patent No. 6355457  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/567,969  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-567-969-5

APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/335,409  
CURRENT FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-335-409-5

Query Match 2.9%; Score 113; DB 3; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2.6;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSWHIKOLYLQHRVLDQMRSGMTVPVLPAPAGHVPEAVTRVFPQVNVTKM-GSWGHFN 272  
DB 4805 IPAEWPLQGVIIHAAGALDD-----GVLDEQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLLAPEDPIPIIGSLFLRELKFGTDHIYGADTFNEMQPPSPSEPSYLAATT 332  
DB 4847 ---NLHETAGNDLAFVLFSSMSGLLSGAGSQSNYAAANTFLDALAAHRAEGLAAQSL 4902

QY 333 A-----VYEAMTAVDTAEAVWLLQGWLFQHPQFQWGPQAIKRAVLG---AVPRGRL--L 379  
DB 4903 AWGPNWDGGMAGLSAA-----LQARLARHGMGALSAPQGTALLGQALARPETOLGAM 4955

QY 380 VLDLFAESQPVYTRTASFQGFPIWCMHLNFGNGHGLFGALEAVNGGPEAARLPNSTMV 439  
DB 4956 SLDVRAASQ-----ASGAAPVPMVRALVRAEARHTAAGAAGAL-----AARL----- 4997

QY 440 GTGMAPEGISONEVYSLMAE-----LQWR-----KDPVPDL-----AAWTSFAARR 482  
DB 4998 --GALPEARRADEVKRVQAEIARVLSWSAASAVPVDRLSDGLDLSLTAVELRNVLGR 5055

QY 483 YGVSS-----HPDAGAAWRLLLRSVYNCSEACRGNHRSPLV-----RRP 521  
DB 5056 VGATLPATLAFDHTVDALTRWLLDKVLAVAEPSVSSAKSSPQVALDEPIAIGICRFP 5115

QY 522 SLOMNTSIWNRSDVEAEWRLLTSAPSLATSPAFRYDLDLTLRQAVQLVSLYE---E 578  
DB 5116 G-----GVADPEFNRLLLEEGSDAVVEPHERWD-----IDAFYDPPDP 5154

QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSMLEQARAA 628  
DB 5155 VRGKMTI-----RFGGFLSDIDRDPAPFGISPREATTMDPQORLLLETSWEAPEFAG 5207

QY 629 AVSE 632  
DB 5208 ILPE 5211

RESULT 5  
US-09-568-102-5  
Sequence 5, Application US/09568102  
Patent No. 6346404  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,102  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30

```
Query Match          2.9%; Score 113; DB 4; Length 7257;
Best Local Similarity 22.3%; Pred. No. 2.6; Mismatches 183; Indels 142; Gaps 22;
Matches 108; Conservative 51;

QY 214 LPPSWHIKOLYQHRVLDQMRSGMTPVLPAPAGHVPEAVTRVFPQVNVTKM-GSWGHN 272
Db 4805 IPAEPWLOGVTHAAGALDD-----GVLDQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLAPEDPFIPIIGSLFLRELKEFGTDHIYGADTFNEMOPSPSPSYLAAAT 332
Db 4847 ---NLHELTAGNDLFAFFVLFS-SMSCLLSGAGSQAQNTFLDAAHRAEGLAAQSL 4902

QY 333 A-----VYBAMTAVDTEAVWLLQGWLFQHOPOFGWPAQIRAVLG---AVPRGL--L 379
Db 4903 ANGPWSGDGMAAGLSAA-----LQARLARHGMGALSQAQNTALLGQALARPETQLGAM 4955

QY 380 VLDLFAESOPVYTRTASFOGQPFIMCLHNFNGHGLFGALEAVNGGPPAARLFPNSTMV 439
Db 4956 SLDVRAASQ-----ASGAAPVPMRALVRAEARTHAAGAOGAL-----AARL----- 4997

QY 440 GTGMAPEGISQNEVYVSLMAE---LGWR-----KDPVPL-----AAWVTSFAARR 482
Db 4998 --GALPEARARRADEVRKVQAEIARVLSWSAASAVPVDRLPSDLGLDSLTAVELRNVLGQR 5055

QY 483 YGVS-----HPDAGAAWRLLLSRVYNCSCGACRGNRSLV-----RRP 521
Db 5056 VCATLPATLAFDHPVDALTRWLLDKVLAVAPSPSSAKSSQVALDEPIAIIIGICRPP 5115

QY 522 SLOMNTSIWYNSRSDVFEAWRLLTSAPSLSATSPAFRYDLDLTLQAOVQLVSLYYE---E 578
Db 5116 G-----GVADPESEFWRLLEEGSDAVVEVPHERWD-----IDAFYDPPDP 5154

QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSWLEQARAA 628
Db 5155 VRGKMT-----RFGGLSDIDRDPFAFGISPREATTMDPQORLLLETSSWEAFERAG 5207

QY 629 AVSE 632
Db 5208 ILPE 5211

RESULT 8
US-09-568-486-5
; Sequence 5, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-5

Query Match          2.9%; Score 113; DB 4; Length 7257;
Best Local Similarity 22.3%; Pred. No. 2.6;
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSWHIKOLYQHRVLDQMRSGMTPVLPAPAGHVPEAVTRVFPQVNVTKM-GSWGHN 272
Db 4805 IPAEPWLOGVTHAAGALDD-----GVLDQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLAPEDPFIPIIGSLFLRELKEFGTDHIYGADTFNEMOPSPSPSYLAAAT 332
Db 4847 ---NLHELTAGNDLFAFFVLFS-SMSCLLSGAGSQAQNTFLDAAHRAEGLAAQSL 4902

QY 333 A-----VYBAMTAVDTEAVWLLQGWLFQHOPOFGWPAQIRAVLG---AVPRGL--L 379
Db 4903 ANGPWSGDGMAAGLSAA-----LQARLARHGMGALSQAQNTALLGQALARPETQLGAM 4955

QY 380 VLDLFAESOPVYTRTASFOGQPFIMCLHNFNGHGLFGALEAVNGGPPAARLFPNSTMV 439
Db 4956 SLDVRAASQ-----ASGAAPVPMRALVRAEARTHAAGAOGAL-----AARL----- 4997

QY 440 GTGMAPEGISQNEVYVSLMAE---LGWR-----KDPVPL-----AAWVTSFAARR 482
Db 4998 --GALPEARARRADEVRKVQAEIARVLSWSAASAVPVDRLPSDLGLDSLTAVELRNVLGQR 5055

QY 483 YGVS-----HPDAGAAWRLLLSRVYNCSCGACRGNRSLV-----RRP 521
Db 5056 VCATLPATLAFDHPVDALTRWLLDKVLAVAPSPSSAKSSQVALDEPIAIIIGICRPP 5115

QY 522 SLOMNTSIWYNSRSDVFEAWRLLTSAPSLSATSPAFRYDLDLTLQAOVQLVSLYYE---E 578
Db 5116 G-----GVADPESEFWRLLEEGSDAVVEVPHERWD-----IDAFYDPPDP 5154

QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSWLEQARAA 628
Db 5155 VRGKMT-----RFGGLSDIDRDPFAFGISPREATTMDPQORLLLETSSWEAFERAG 5207

QY 629 AVSE 632
Db 5208 ILPE 5211

RESULT 7
US-09-568-480-5
; Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-5

Query Match          2.9%; Score 113; DB 4; Length 7257;
Best Local Similarity 22.3%; Pred. No. 2.6;
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;
```

Db 4847 ---NLHETAGNDLAFVFLFS-SMSGLLGSAGQSNYAANTFLDALAAHRRRAEGLAAQSL 4902  
QY 333 A-----VYEAMTAVDTAEVWMLLQGWLFQHQPOFWGPAQIRAVLG---AVPRGRL--L 379  
Db 4903 AWGPNWDGMAAGLSAA-----LQARLARHGMGALSQAQGTALLGQALARPETQLGAM 4955  
QY 380 VLDLFAESQPVYTRTASTFQOGQFFIWCMLHNFNGHGLFGALEAVNGGPEAARLFPNSTMV 439  
Db 4956 SLDVRAASQ-----ASGAAPVPPVWRALVRAEARHTAAGAOGAL-----AARL----- 4997  
QY 440 GTGMAPEGISQNEVVYSLMAE-----LGWR-----KDPVDDL-----AAWVTSAARR 482  
Db 4998 --GALPEARRADEVRKVVQAEIARVLSMSAASAVPVDRLPLSLDGLDLSLTAVELRNVLGQR 5055  
QY 483 YGVY-----HPDAGAAWRLLLRVYNCSEACRGNRSPV-----RRP 521  
Db 5056 VGATLPATLAFPHPTVDALTRWLLDKVLAVAEPSVSSAKSSQVVALDEPIAIIIGICREP 5115  
QY 522 SLOMNTSIWYNSRDVFEAWRLLLTSAPSLSATSPAFRYDLDLTLTRQAVQELVSLYE---E 578  
Db 5116 G-----GVADPESEFWRLLEEGSDAVVEVPHERWD-----IDAFYDPPDP 5154  
QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSLWEQARAA 628  
Db 5155 VRGKWT-----RFGGFLSDIDRFPDPAFFGISPREATTMDPQORLLLETSSWEAFERAG 5207  
QY 629 AVSE 632  
Db 5208 ILPE 5211  
RESULT 9  
US-09-568-472-5  
; Sequence 5, Application US/09568472  
; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,472  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 7257  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-472-5

Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2,6;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;  
QY 214 LPSPWHIKOLYQHRVLDQMRSGMTFVLPAFAGHVPEAVTRVFPQVNTKMG-SWGHFN 272  
Db 4805 IPAENPLQGVIIHAAGALDD-----GVLDEQTTDRFSRVLPKVTGAW---- 4846  
QY 273 CSVSCSFLLAPDPIPIITGLFLRELKEFGFDHYIGADTFNEMQPPSEPSYLAANA 332  
Db 4847 ---NLHETAGNDLAFVFLFS-SMSGLLGSAGQSNYAANTFLDALAAHRRRAEGLAAQSL 4902  
QY 333 A-----VYEAMTAVDTAEVWMLLQGWLFQHQPOFWGPAQIRAVLG---AVPRGRL--L 379  
Db 4903 AWGPNWDGMAAGLSAA-----LQARLARHGMGALSQAQGTALLGQALARPETQLGAM 4955

QY 380 VLDLFAESQPVYTRTASTFQOGQFFIWCMLHNFNGHGLFGALEAVNGGPEAARLFPNSTMV 439  
Db 4956 SLDVRAASQ-----ASGAAPVPPVWRALVRAEARHTAAGAOGAL-----AARL----- 4997  
QY 440 GTGMAPEGISQNEVVYSLMAE-----LGWR-----KDPVDDL-----AAWVTSAARR 482  
Db 4998 --GALPEARRADEVRKVVQAEIARVLSMSAASAVPVDRLPLSLDGLDLSLTAVELRNVLGQR 5055  
QY 483 YGVY-----HPDAGAAWRLLLRVYNCSEACRGNRSPV-----RRP 521  
Db 5056 VGATLPATLAFPHPTVDALTRWLLDKVLAVAEPSVSSAKSSQVVALDEPIAIIIGICREP 5115  
QY 522 SLOMNTSIWYNSRDVFEAWRLLLTSAPSLSATSPAFRYDLDLTLTRQAVQELVSLYE---E 578  
Db 5116 G-----GVADPESEFWRLLEEGSDAVVEVPHERWD-----IDAFYDPPDP 5154  
QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSLWEQARAA 628  
Db 5155 VRGKWT-----RFGGFLSDIDRFPDPAFFGISPREATTMDPQORLLLETSSWEAFERAG 5207  
QY 629 AVSE 632  
Db 5208 ILPE 5211  
RESULT 10  
US-08-972-927-3  
; Sequence 3, Application US/08972927  
; Patent No. 6166290  
; GENERAL INFORMATION:  
; APPLICANT: Rea, Phillip A  
; APPLICANT: Lu, Yu-Ping  
; APPLICANT: Li, Ze-Sheng  
; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN  
; TITLE OF INVENTION: PLANTS  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C., 22nd  
; STREET: One Commerce Square, 2005 Market Street, 22nd  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: US  
; ZIP: 19103-7086  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/972,927  
; FILING DATE: 18-NOV-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/031,040  
; FILING DATE: 18-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/061,328  
; FILING DATE: 08-OCT-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Doyle Leary Ph.D., Kathryn  
; REGISTRATION NUMBER: 36,317  
; REFERENCE/DOCKET NUMBER: 9596-1202  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-965-1284  
; TELEFAX: 215-567-2991  
; TELEX: 831-494  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1621 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-972-927-3

Query Match 2.7%; Score 107.5; DB 4; Length 1621;  
Best Local Similarity 19.8%; Pred. No. 0.72;  
Matches 126; Conservative 75; Mismatches 214; Indels 221; Gaps 33;

QY 175 WQVYLALG---LQAEINEEFTGPAFLAWGRMGNL-----HTWDGPL--PPSHHIKOL 223  
DB 241 WNPMLTGLSKRPLTKDQVYVYLDT-----WDQETLFTSFQHSWDKELQKQPQWLLRAL 294  
QY 224 YLQHRVLDQMRSGMTPLVLPAPAGHVPEAVTRVFQVNVTKMGs---WGHF--NCYSVCSF 279  
DB 295 N-----NSLGGRFWGGFWKIGNDCS- 315  
QY 280 LLAPEDPIPIIGSLFLRLIIEFGTD-----HIYGADTFNEMQPPSSPSFYLA 328  
DB 316 -----QFVGPLLLNQLKSMQEDAPAMGYIYAFSIFGVVFGVL---CEAQYFQ 362  
QY 329 AATTAVYEAMTAVDEAVWLLQGWLFQHQPFQWGPQAIKRAVLGAVPRGRLLVLDLFAES- 387  
DB 363 NVNRVGYRLRSA-----LIAVFRKSLRTNNGRRKFQTKIT--NLMTD--AESL 410  
QY 388 QPVYTRTASFQOGPF-----IWCMHLNFGNGHGLFGALEAVNGGPEAARLFPNSTMVGTGM 443  
DB 411 QOICOSLHTMWSAPFRILIALILLYQQLGVASLIGALLLV-----LMFPLQTVLIISKM 463  
QY 444 ---APEGISONEVYSLMAELGWRKDPVDPDLAAVTSFPAARYGVSHPDAGAAWRLLLRS 500  
DB 464 QKLTKEGQRTDKRIGLNMNEVLAAAMDTV--KCYAWENSE----- 500  
QY 501 VYNCSEACRGNHNSPLVRPRLSLOMNTSIWYNSRSDVFEAWRL--LLTSAPSLATSPAFR-Y 558  
DB 501 -----QSKVQTVRDELS-----WFKSOLLGALNMFILNSIPLVTVISFGVF 544  
QY 559 DLL--DLTRQAVOELVSLYIEARSAYLSKELASLLR-----AGGVLAYELLPAL 606  
DB 545 TLGGDLT-----PAR-AFTSLSLFAVLRFPLEMPLNIIQVNVANVSLKRL 590  
QY 607 DEVLASDRFLGSLWLEQARAAVSEAEADFY-----EQNSRYQLTWCPEGNILDYANK 561  
DB 591 EEVLATEERILLPNPIEPGEPAISIRNGYFSWDSKGDRTLSNINLNDVPLGSLVAVVGS 650  
QY 662 -----QLAGLVANYTPR--WRLFLBALVDSVDAQGIPFQOH 695  
DB 651 TGEKTSLSIALGELPATSDAIVTLRGSVA--YVPQVSW-IFNATVRDNLILFGSPFDEE 707  
QY 696 QFDK--NVFOLQAFVLSKQRYPSQPRGDTVDLAKK 729  
DB 708 KYERAIDVTSKHDLEL-----LPGGDLTEIGER 736

RESULT 11  
US-08-972-927-6  
Sequence 6, Application US/08972927  
Patent No. 6166290  
GENERAL INFORMATION:  
APPLICANT: Rea, Philip A  
APPLICANT: Lu, Yu-Ping  
APPLICANT: Li, Ze-Sheng  
TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN PLANTS  
TITLE OF INVENTION: PLANTS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
STREET: One Commerce Square, 2005 Market Street, 22nd  
STREET: Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: US  
ZIP: 19103-7086  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/972,927  
FILING DATE: 18-NOV-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,040  
FILING DATE: 18-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/061,328  
FILING DATE: 08-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Doyle Leary Ph.D., Kathryn  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: 9596-12U2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-965-1284  
TELEFAX: 215-567-2991  
TELEX: 831-494  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1622 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-972-927-6

Query Match 2.7%; Score 107.5; DB 4; Length 1622;  
Best Local Similarity 21.1%; Pred. No. 0.72;  
Matches 116; Conservative 63; Mismatches 167; Indels 203; Gaps 30;

QY 268 WGHF--NCYSVCSFLLAPEDPIPIIGSLFLRLIIEFGTDHIYGADTFNEMQPPSSPSY 326  
DB 304 WGGFWKIGNDCS-----QFVGPLLLNELLK-----SMQ--LNEPAW 337  
QY 327 LAAATAVYEAMTAVDEAVWLLQGWLFQHQPFQWGPQAIKRAVLGAVPR-----GR- 377  
DB 338 IG-----YIYAFSIFGVVGLGVLCSEAQYFQNVNRV--GYRLRSALIAAVFRKSLRLTNEGRK 392  
QY 378 -----LLVLDLFAES--QPVYTRTASFQOGPF-----IWCMHLNFGNGHGLFGALEAV 423  
DB 393 KEQTKINLMTD--AESLQOICOSLHTMWSAPFRILIALILLYQQLGVASLIGALLLV 450  
QY 424 NGGPEAARLFPNSTMV---GTGMAPEGISQNEVYSLMAELGWRKDPVDPDLAAVTSFAA 480  
DB 451 -----LMFPIQTVIISKTKLTKEGQLQRTDKRIGLNMNEVLAAAMDTV--KCYAWENSEF-- 500  
QY 481 RYGVSHPDAGAAWRLLLRSVYNCSEACRGNHNSPLVRPRLSLOMNTSIWYNSRSDVFEAW 540  
DB 501 -----QSKVQTVRDELS-----WFKKAQLLSAF 524  
QY 541 RL--LLTSAPSLATSPAFR-YDLL--DLTRQAVOELVSLYIEARSAYLSKELASLLR--- 593  
DB 525 NMFILNSIPLVTVVVSFGVSLGGDLT-----PAR-AFTSLSLFVSIRPL 570  
QY 594 -----AGGVLAYELLPALDEVLASDSRFLG-----SWLQARAA 628  
DB 571 FMLPNIITQMVNANVSLNRLEVLSTEEVLLPNPIEPGQPAISIRNGYFSWDSKADRP 630  
QY 629 AVSEAEADFYEQNSRYQLTWCPEGNIL-----DYANKOLA 664  
DB 631 TLSNINLDI-----PLGSLVAVVSGTGEKTSLSIALGELPARSDATVTLR 677  
QY 665 GLVANYTPR--WRLFLBALVDSVDAQGIPFQOHQFDK--NVFOLQAFVLSKQRYPSOPR 720  
DB 678 GSAV--YVPQVSW-IFNATVRDNLILFGAFDQEKYERVIDTALQHDLEL-----LPG 727  
QY 721 GDTVDLAKK 729





Db 3457 DAEALHTVLPALSS-----WRRRRVHRRLODWRYRVE---WKPPPAALDEVILGGWLFVVP 3510  
QY 375 RG-----RLLVLDLFAESQPVYTRTASFCQPFIMWMLHNFQGN 413  
Db 3511 RGLADGVARVVAANTARGGEVSVVELDPTDRDRAYAEAVAGRGVSVFSLWDDRR 3570  
QY 414 HGLFGALEAVNGGPEARLPNSTVGTGMAPEG-----ISONEVVYSLMAELGWRKDPV 468  
Db 3571 HSEHSVVA-----GLAASLVLAQALVDLGRVGGPRLWLVTGCAVAG-PSDAGVVVDV 3625  
QY 469 PDLAANVTFAARRYGVSHPDAGAAWRLLRSVYNGSGEACR-----GHNRSPLVR 519  
Db 3626 ---QAQWGF-GRVLGLEHPE---LWGLVDLPVGVDEEVCRRFVGWASAGFEDQAVR 3678  
QY 520 RPSLOMNTSIWYNR 533  
Db 3679 -----GSGVWYVR 3686

RESULT 14  
US-08-313-185-49  
; Sequence 49, Application US/08313185  
; Patent No. 5851763  
; GENERAL INFORMATION:  
; APPLICANT: Heym, Beate  
; APPLICANT: Cole, Stewart  
; APPLICANT: Young, Douglas  
; APPLICANT: Zhang, Ying  
; APPLICANT: Honore, Nadine  
; APPLICANT: Telenti, Amalio  
; APPLICANT: Bodmer, Thomas  
; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance  
; TITLE OF INVENTION: in Mycobacterium Tuberculosis  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 12-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 02356.0068-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-313-185-49

Query Match 2.7%; Score 104.5; DB 2; Length 726;  
Best Local Similarity 18.3%; Pred. No. 0.36;  
Matches 125; Conservative 66; Mismatches 189; Indels 303; Gaps 32;  
QY 86 AAGLHRYLDFCGCHVANGSQLRLPRPLPAVPGELTEATNR-----YR----- 130

Db 26 SAGAGTTTRD-----WPNQLRV-----DLLNQHNSNRNPLGEDFDYRKESKL 69  
QY 131 -YY-----QNYCTQSYSFVWMDWARWEREIDWMLALNLAWS----- 169  
Db 70 DYYGLKDLKALLTESQP--WN-----PADWGSVAGLFIRMAWHGAGTYSIDRGGA 120  
QY 170 -----GOEAIWORYVLAGLTOAEINEFFTG 196  
Db 121 GRGOORFAPLNSWPDNVDLKARRLLWPIKQYKOKISWADLFIAGNVALENSGFT-- 178  
QY 197 AFLANGMGNLHTDGPPLPPSWHIIKOLYLOHRLVQDMRSGMTPLVPAPAGHVPEAVTR 256  
Db 179 --FGFG-AGREDVWEPDLVDNMGDEKAWLTHR-----H-PEALAKA 215  
QY 257 FPOVNVTKMGSMGFHCYSFCSFLLAPEDPIFPIIGSLFLRELIFKEFGTDHYGADTFNE 316  
Db 216 --PLGATEMG-----LIYVNEPGP-----DH----- 234  
QY 317 MQPPSSERSYLAATYAYEAMTAVDTAVWLLQGLWLFQHQPFQWGPQAIQIRAVLGAVPRG 376  
Db 235 -----SGEPLSAAAIRATFGNMGMDDETVALIAGHTLKGTHGAGPT---SNVGPDP-- 285  
QY 377 RLLVLDLFAESQPVYTR-----TASFQGPFIWMLHNFQGNH---GLF-- 417  
Db 286 -----EAAPIEEQGLGWASTYSGVGADAITSGLEVWVTQTPTQWNSYFFENLFKY 336  
QY 418 -----GAL--EAYNGGPEEARLPNSTVMGTGMAPEGISQNEVVYSLMAELGWRKD 466  
Db 337 EWWQTRSPAGAIQFEAV-----DAPEIIPD-----PFDPSSKKRKPMTLVTDLTRFD 383  
QY 467 P-----VPDLAAWTSFAARYGYVSHPDAGAAWRLLRSVYNGSGEACRGNRSP 517  
Db 384 PEFKISRRLNDPOAFNEAFARAWFKLTHRDMGPKSYI----- 423  
QY 518 VRPSLOMNTSIWYNRSDVFEAWRLLLTASPLATSAFRRYDLDL-----TROAVOEL 571  
Db 424 --GPEVPKEDLIQD-----PLQPIYNPT-EQDIIDLKFAIADSGLSVSEL 467  
QY 572 VSLYYEARSAYLSKELASLLRAG-----GVLAYELLPALDEV 610  
Db 468 VSV-----AWAS---ASTFRGDKRGANGARLALMPORDWVYNAARALPVLEKIQ 517  
QY 611 ASDSRFLGSLWLEQARAAVSEA 633  
Db 518 KESGKASLADIIVLAGVVGVEKA 540

RESULT 15  
US-08-459-499-13  
; Sequence 13, Application US/08459499  
; Patent No. 5871912  
; GENERAL INFORMATION:  
; APPLICANT: Heym, Beate  
; APPLICANT: Cole, Stewart T.  
; APPLICANT: Young, Douglas B.  
; APPLICANT: Zhang, Ying  
; TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods  
; TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isoni  
; TITLE OF INVENTION: Amended  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.3  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,499  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/875,940  
FILING DATE: 30-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/929,206  
FILING DATE: 27-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/029,655  
FILING DATE: 11-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0110-03000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-459-499-13

Query Match 2.7%; Score 104.5; DB 2; Length 726;  
Best Local Similarity 18.3%; Pred. No. 0.36;  
Matches 125; Conservative 66; Mismatches 189; Indels 303; Gaps 32;

|    |     |  |     |
|----|-----|--|-----|
| QY | 86  | _AGLHRYLRDFCCCHVAVSGSQLRPLRPAPVPGELTEATPNR-----YR-----       | 130 |
| DB | 26  | SAGAGTTTRD-----WMPNQLRV-----DLLNQHSNRNPLGDEDFYRKEFSKL        | 69  |
| QY | 131 | -YY-----QNVCTQSYFVWMDWARWEREIDMMLNGINLALAWS-----             | 169 |
| DB | 70  | DYGLKDKLKALTESQP--WW-----PADWGSYAGLFIRMAWHGAGTYRSIDGRGA      | 120 |
| QY | 170 | -----GQBAIWQRYVYALGLTQAEINEFFTGP                             | 196 |
| DB | 121 | GRGQQRFAPLNSWPDNVSLDKARRLLWPIKQYQKISWADLFILAGNVALENSGERT--   | 178 |
| QY | 197 | AFLAWGRGNLHTWDGFLPPSHIKQLYLQHRVLDQMRSGFTPLVPAPAGHVPEAVTRV    | 256 |
| DB | 179 | --FGFG--AGREDVMEPDLVDNMGDEKAWLTHR-----H--PEALAKA             | 215 |
| QY | 257 | FPOVNTKMGSMGHFNCYSYSCSFLLAPEDPFPPIIGSLFLRELKKEFGTDHIYGADTFNE | 316 |
| DB | 216 | --PLGATEMG-----LIYVNEGP-----DH-----                          | 234 |
| QY | 317 | MQPPSEPSVLAATAVYEAATVAVTEAVMLQGNLFQHQPOFGWPAQIRAVLGAVPRG     | 376 |
| DB | 235 | -----SGEPLSAAAIARATFGNMGMDDETVALIAGGHTLGKTHGAGPT---SNVGPDP-- | 285 |
| QY | 377 | RLVLVDLFAESQPYVTR-----TASFQGPFIWCMHLNFGGNH---GLF--           | 417 |
| DB | 286 | -----EAPFIEGGLWAGTYSGVGCADAITSGLEVVTQTPTQWSNYFFENLFKY        | 336 |
| QY | 418 | -----GAL--EAVNGGPEAARLPNSTWVGTGMAPEGISQNEVYVSLMAELGWRKD      | 466 |
| DB | 337 | EWQTRSPACAIQFEAV-----DAPEIIPD-----PEDPSKKRKPMTMLVTDLTFRD     | 383 |
| QY | 467 | P-----VPDLAAVWTSFAARRYGVSHPDAGAAWRLLLRVYNCSGEACRGNRSPL       | 517 |
| DB | 384 | PEFEKISRRLNDPQAFNEAFARAWFKLTHRDMGPKSYI-----                  | 423 |
| QY | 518 | VRRPSLQMNTSIWYNRSDVFEAWRLLLTSAPSLATSPAFRYDLDDL-----TRQAVQEL  | 571 |
| DB | 424 | --GPEVPEKEDLIWQD-----PLPQPIYNPT--EQDIIDLKFAIDAGLSVSEL        | 467 |

|    |     |  |     |
|----|-----|--|-----|
| QY | 572 | VSLYYEEARSAYLSKELASLLRAG-----GVLAYELLPALDEVL                 | 610 |
| DB | 468 | VSV-----AWAS-----ASTFRGGDKRGANGARLALMPQRDWDVYNAAAVRALPVLEKIQ | 517 |
| QY | 611 | ASDSRFLILGSLWLEQARAAAVSEA                                    | 633 |
| DB | 518 | KESGKASLADIIVLAGVVGVEKA                                      | 540 |

Search completed: September 21, 2002, 12:24:27  
Job time: 42 sec

| Result No. | Query Match % | Score  | Length | DB | ID     | Description        |
|------------|---------------|--------|--------|----|--------|--------------------|
|            |               |        |        |    |        |                    |
| 1          | 100.0         | 3939   | 743    | 1  | G02370 | alpha-N-acetylgluc |
| 2          | 26.9          | 1060.5 | 770    | 2  | C87316 | alpha-N-acetylgluc |
| 3          | 3.1           | 123.5  | 497    | 1  | G02938 | probable debrisoqu |
| 4          | 121           | 121    | 4342   | 2  | H83343 | probable non-ribos |
| 5          | 3.0           | 3649   | 31     | 1  | S18268 | Delta-(L-alpha-ami |
| 6          | 117.5         | 3.0    | 497    | 1  | O4HUD1 | debrisoquine 4-hyd |
| 7          | 117           | 3.0    | 1374   | 2  | D72593 | hypothetical prote |
| 8          | 115.5         | 2.9    | 973    | 2  | T35238 | hypothetical prote |
| 9          | 115           | 2.9    | 725    | 2  | C87485 | probable secreted  |
| 10         | 114           | 2.9    | 479    | 2  | T31293 | ComEC/Rec2 family  |
| 11         | 113           | 2.9    | 1281   | 2  | G83405 | aldehyde dehydroge |
| 12         | 112           | 2.8    | 1323   | 2  | T78557 | hypothetical prote |
| 13         | 111           | 2.8    | 1027   | 2  | H87316 | N-methyl-D-aspart  |
| 14         | 109.5         | 2.8    | 285    | 2  | F87285 | hypothetical prote |
| 15         | 109.5         | 2.8    | 814    | 2  | T47641 | conserved hypotet  |
| 16         | 109.5         | 2.8    | 1377   | 2  | C65159 | hypothetical prote |
| 17         | 109.5         | 2.8    | 1411   | 2  | E65145 | rhsA protein precu |
| 18         | 109           | 2.8    | 1080   | 2  | T19048 | rhsB protein precu |
| 19         | 107.5         | 2.7    | 1622   | 2  | D86428 | probable Pro-X car |
| 20         | 107           | 2.7    | 741    | 2  | A83271 | glutathione S-conj |
| 21         | 106.5         | 2.7    | 666    | 2  | A87577 | hypothetical prote |
| 22         | 106.5         | 2.7    | 1678   | 2  | T35547 | oligopeptide trans |
| 23         | 104.5         | 2.7    | 726    | 1  | CSECHP | hypothetical prote |
| 24         | 104.5         | 2.7    | 726    | 2  | G91237 | catalase [BC 1.11. |
| 25         | 104           | 2.6    | 1230   | 2  | T07663 | hydroperoxidase H  |
| 26         | 103.5         | 2.6    | 726    | 2  | C96085 | soluble starch syn |
| 27         | 103.5         | 2.6    | 1737   | 2  | T00209 | catalase, hydrop   |
| 28         | 103           | 2.6    | 698    | 2  | D90771 | MEG8 protein - hu  |
| 29         | 103           | 2.6    | 698    | 2  | H85632 | hypothetical prote |

||||| 361 WGAQIRAVLGVPRGRLLVLDLFAESQPVYTRTASFGQPIWMLHNFNGHGLFGL 420  
Db  
QY 421 EAVNGGPEAARLFPNSTMVGTCMAPEGISQNEVYSLMAELGWRKDPDPDLAAWVTSFAA 480  
|||||  
Db 421 EAVNGGPEAARLFPNSTMVGTCMAPEGISQNEVYSLMAELGWRKDPDPDLAAWVTSFAA 480  
QY 481 RRYGVSHPDAGAAWRLRLLSVYVNCSEACRGNRSPLYRRPSLQWNTSIWYNRSDFEAW 540  
Db 481 RRYGVSHPDAGAAWRLRLLSVYVNCSEACRGNRSPLYRRPSLQWNTSIWYNRSDFEAW 540  
QY 541 RLLTSAPSLSATSPAFRYDLDLTLQAVQELVSLYEEARSAYLSKELASLLRAGGVLAY 600  
Db 541 RLLTSAPSLSATSPAFRYDLDLTLQAVQELVSLYEEARSAYLSKELASLLRAGGVLAY 600  
QY 601 ELLPALDEVLASDSRLLGSLWEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYAN 660  
Db 601 ELLPALDEVLASDSRLLGSLWEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYAN 660  
QY 661 KQLAGLVANYTPRWRFLFLEALVDSVAQGIPTQOHQFQKQVFNQLEQAFVLSKQRYPSQPR 720  
Db 661 KQLAGLVANYTPRWRFLFLEALVDSVAQGIPTQOHQFQKQVFNQLEQAFVLSKQRYPSQPR 720  
QY 721 GDTVDLAKKIFLYKYPGWAGSW 743  
Db 721 GDTVDLAKKIFLYKYPGWAGSW 743  
RESULT 2  
C87316  
alpha-N-acetylglucosaminidase [Imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: C87316  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.J.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Emdolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; UID:21173698; PMID:11259647  
A:Accession: C87316  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-770 <STO>  
A:CROSS-references: GB:AE005673; NID:gl3421729; PIDN:AAK22527.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0540  
C:Superfamily: human alpha-N-acetylglucosaminidase  
Query Match 26.9%; Score 1060.5; DB 2; Length 770;  
Best Local Similarity 33.0%; Pred. No. 1.3e-68;  
Matches 260; Conservative 132; Mismatches 302; Indels 93; Gaps 19;  
QY 2 EAVAVAAVGVLLLAGAGGADEARAAVRAVRLVLLGPGPAADFVSVERALAAKPG 61  
Db 11 QATSLAAAFVSPALAAAGTDG----VAAARASLKRFLGRRLAG-----AHLTVTPG 59  
QY 62 LDT--YSLGGGAARVRVRGSGTVAAAGLHRYLRVDFCGCHVAGWSQLRPLRPAPVG 119  
Db 60 AERSWYAIGKGA--ISIGDSPVLYRGAYAHLRQAGLAHVSGEGRVQAGAVPAGAG 118  
QY 120 ELTEATPNRYVQVNTQSYFVWDMARWERIDMNLNGINLALAWSGQEAIRVRY 179  
Db 119 ARVE--TPFRHAYLNTCYGTYTTPWGWGRWTRIDWNAAGIDMPLAMEGQEVVWRLW 177  
QY 180 LALGLTOAINEFTGPAFLAWGWMGNLHTWDGPLPSWHIKOLYLQHRVLDQMRSGMT 239  
Db 178 REFLGSAELADYFSGPAFTPHWRMGNIETGKAPLPTAWIDKKDLQVKILGRNRSLSMT 237  
QY 240 PVLPAFAGHVPEAVTRVFPVNVNTKMGSGHFNCSYSCSFLLAPEDPIFFIIGSLFUREL 299  
Db 238 PILPAFGGYVPKAFENKPNKAIYRMRPWEGFHETY----WLDPADPLFAKIAARFLALY 293

QY 300 IKEFTDHIYGADTFNEMQPP-----SSEPSY-----LA 328  
Db 294 TETGAGCTYGLADSENEMLPPIINAGDARDADAAYGDTANTAVTKKVEVDPAKQAORLA 353  
QY 329 AATTAVYEMTAVDTEAVWLQGLWLFHQHQPQFWGPAQITRAVLGAVPRGRLLVLDLFAESQ 388  
Db 354 AVGKAIYDSIRQTRDPAVVMQWGLFGADSHFMDPAATISAYLSLVPDDKMLIDIGNDRY 413  
QY 389 P-VYTRTASFQGPPIWMLHNFNGHGLFGL-----EAVNGGPEAARLFPNSTMVG 440  
Db 414 PNWKNKAFGKPMIYGYVHNYGSGNPVYGDGLFYRQDIPAIANPDAGKL-----AG 467  
QY 441 TGMAGEGISQNEVYSLMAELGWRKDPDPDLAAWVTSFAARRYGVSHPDAGAAWRLRLLS 500  
Db 468 GCMFEGELHNSIYVYEAAYDLAWSEGOASP-ATWLTTRYARARYGKTSPLDAALQOLVEA 526  
QY 501 VYNCSEACRGNRSP-----LVRRPSLQWNTSIWY--NRSDVFEAWRLRLLSAPSLSATSP 554  
Db 527 AFSTRYSPRWKSKAGAYLFFKRPRTATVGPDPQHPGDRAKLEAAVAKALTALAPTYGOEP 586  
QY 555 AFPRYDLDLTLQAVQELVSLYEEARSAYLSKELASLLRAGGVLAYELLPALDEVLASDS 614  
Db 587 LEVLDLTATRLHATMKIDDLLOVAVAYRGGDTA----AGDAARVEI-----EALASI 637  
QY 615 RFLG-----SWLEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYANKQLAGLV 667  
Db 638 DKLLGVQPDTLATWIDEARAYGDTPADAAAYVANAQAQVTWGGEGNLDYASKAWQGLY 697  
QY 668 ANYTTPRWRFLFLEALVDSVAQGIPTQOHQFQKQVFNQLEQAFVLSKORY-----PSQPRGDT 723  
Db 698 KSFYLPWRMSRFLDAL--KAACTGTDEVTVTRGGVAVARAWVEAEVAYRREKPADPIGEI 755  
QY 724 VDLAKKI 730  
Db 756 KTLIARI 762  
RESULT 3  
G02938  
probable debrisquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 - crab-eating maca  
C:Species: Macaca fascicularis (crab-eating macaque)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Mar-2000  
C:Accession: G02938  
R:Lawton, M.P.; Laddison, K.J.; Speirs, A.A.; Mankowski, D.C.; Tweedie, D.J.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: G12616  
A:Accession: G02938  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-497 <LAW>  
A:CROSS-references: EMBL:U38218; NID:gl022899; PIDN:AAK79722.1; PID:gl022900  
C:Genetics:  
A:Gene: CYP2D17  
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein; monooxygenase; oxidoreductase;  
F:302-465/Domain: cytochrome P450 homology <P45>  
F:443/Binding site: heme iron (Cys) (axial ligand) #status predicted  
Query Match 3.1%; Score 123.5; DB 1; Length 497;  
Best Local Similarity 22.9%; Pred. No. 0.45;  
Matches 104; Conservative 47; Mismatches 140; Indels 163; Gaps 22;  
QY 343 TEAWLL-----QGWLFQHQPFQWGAQITRAVLGAVPRGRLLVLDLFAESQPVYTRT 394  
Db 12 TVAIFLLVLDLHRRQRWAARYPP--GPLPLPGL-----GNLLHVD----- 50  
QY 395 ASFGQGPFTWCLHNFNGN-----HGLFGALEA-VNGGPEAARL--FPNST 437  
Db 51 --FRNTPYCEDQLRRRFGNVSFQLQAWTPVVLNGLAAVREALVTCGEDTADPPPINQ 108  
QY 438 MVGTGMAPEGI-----SQNEVYSLMAELGWRKDPDPDLAAWVTSFAARRYGVSH 487

```
Db 109 VLGFGRSQGVFLARYGPANRQRRESVSLRNILGLK---KLEQWVTEEAACLCAAFT 165
Qy 488 PDAGAAWR---LLLRVSYNCSGEACRHNRSPLVRPSLOMNTSIWYNRSDVFEAWRLLL 544
Db 166 DQAGRFRPNSLDKAVSN-----VTASLTYGR----- 193
Qy 545 TSAPSLATSPAFRYD-----LLDLTRQAVQELVSLYYBEARSAYLSKELAS---LLR-- 593
Db 194 -----RFEYDDPRFLRFLDLTHEALKE-----ESGFUREVNLNAPLLLRIP 234
Qy 594 --AGGVLAWE--LLPALDEVL-----ADSRELLGSWLEQARAAAVSEAEADFYBQ 640
Db 235 GLAGKVLRSOKAFLOTDELLETHRMTWPAQPPRDLTEAFLAEMEKAR-GNPESSEFNEE 293
Qy 641 NSRYQLTLMGPEGNILDYANKLAGLVANVYTPRWRLFLLEALVDSVAQGIPIQQOHQFDKN 700
Db 294 NLR-----MVVADLSAGNVTTSTLAWGLLLMLHPDVQRRV---QQEIDDV 338
Qy 701 VFQLEQAFLVSKORYP-----SOPRGDTVDL 726
Db 339 IGOVRRPEMGDQARMPYTTAVIHEVQRFGDIVPL 372
```

RESULT 4

H83343  
probable non-ribosomal peptide synthetase PA2424 [Imported] - Pseudomonas aeruginosa (st  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: H83343  
R;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A;Reference number: A82950; MUID:20437337  
A;Accession: H83343  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-4342 <STO>  
A;Cross-references: GB:A8004669; GB:A8004091; NID:99948460; PIDN:AAG05812.1; GSPDB:GN001  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA2424  
C;Superfamily: acyl carrier protein homology; acetate--CoA ligase homology  
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F:60-553/Domain: acetate--CoA ligase homology <ACLI>  
F:584-652/Domain: acyl carrier protein homology <ACP1>  
F:1174-1622/Domain: acetate--CoA ligase homology <ACLI2>  
F:1637-1705/Domain: acyl carrier protein homology <ACP2>  
F:2232-2689/Domain: acetate--CoA ligase homology <ACLI3>  
F:2706-2773/Domain: acetate--CoA ligase homology <ACLI4>  
F:3759-4230/Domain: acyl carrier protein homology <ACP3>  
F:4248-4316/Domain: acyl carrier protein homology <ACP4>  
F:1669,2738,4280/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 3.18; Score 121; DB 2; Length 4342;  
Best Local Similarity 21.4%; Pred. No. 14;  
Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;

```
Qy 4 VAVAA-----AVGVLLLAGAGGA-----AGDEARCAAAVRA-LVARLL 40
Db 1181 VAICERSPOLLVGLLAIVRAGGAYVLPDDYPSERLAYMLADSGVELLLTQAHFLERL- 1239
Qy 41 GPG-----PAADFVSVERALAAKPGLDYSLGGGGAARVRGSGTGVAAAGLHRYLRD 95
Db 1240 -PGAEGVTPICLDSKLDNNPQAPGLHLH--GDNLAIVYTSGSTQCPKGVG----- 1289
Qy 96 FCGCHVAVSGSOLRPLRPFAVPGCE--LTATPNRYRYQNVCTQSYFVWMDWARWERE 153
Db 1290 --NTHAALA-ERLQMMQATVTLGDGDDVLMQKAP-----VSFDSVSW--E 1328
```

```
Qy 154 IDWMLANGINLALWSGQ-----EAIWQVYLLALGLTQAEINEFFTGPAFLAWGRMG 205
Db 1329 CFWPLVTGCRVLVAAPGEHRDPARLVELVRQFGVITLHFVPPLLQLFIDEPGVAACGSILR 1388
Qy 206 NLHTWDGGLPPSPSHLKLQLYLQHRVLQDWRSGMTVPVLPFAFAGHVPVAVTRVPQVNVTKM 265
Db 1389 RLFSGEALPAE-----LRNRVLR-----LPAVALHNRYGTET--AINVT-- 1428
Qy 266 GSWGHFNGSYSGSFLLAPED-----PIFFLIGSLFLURELIKFGTTHIYGADTFENEM--- 317
Db 1429 -----HWQC-----RAEDGERSPIGRPLGNVVCVLDAAEF---NLLPAGVAGELCIGG 1473
Qy 318 -----QPPSSPSPLAAATTAAYEAMTAVDTFVAVLQGWL-----FQHQPFQWG- 362
Db 1474 LGLARGYLGRPALSAERFVADPFSAAGERLYRTGDRARNADGVLEYLGRLDQOVKLRGF 1533
Qy 363 ---PAQIRAVILGAVP---RGRLLVLDLPAESQPV---YFRTASFGQGFPIWMLHNFQGNH 414
Db 1534 RIEPEIQARLLAOPGVAQAVVIREGVAGSOLVGYTYTGAVGAAEA-----EQNQ 1584
Qy 415 GLFGALEAVNGGPEAARLFPNSTMVGTGMAPEGISQNEVVYSLMAELGWRK---DPVPD 470
Db 1585 RLRAALQA-----ELPEYMVPTQLMRLAQMPLGPS-GKLDTRALPEPVWQQRHEHVRTE 1638
Qy 471 L-----AAWVTSFAARYG-----VSHPDAGAANRLLLRSYNCS-- 505
Db 1639 LQRTIAIWSVGLPRVGLRDXDDFFELGGHSLLATRIIVSRTRQACDVLPURLFAESEL 1698
Qy 506 -----GEACRHNRSPLVRPSLQMNTSIWYNRSDVFEAWRLLLTSAPSLATSPAF 556
Db 1699 EAFCEQVRAAAGRTDSHGAIIRIDREOPVPLSYSQRMWFLWQL-----EPD---SPAY 1751
Qy 557 RY-----DLDLTR--QAVQELVSLYYEARSAYLSKELASLLRA---GGV-LAYELL 603
Db 1752 NVGGLARLSCGLDVARFAALQALVQ-RHETLRTTFFSVDPGVQVRVHGDDGLHMDQDF 1810
Qy 604 PALD-----EVLASDRFLGLSLEQA---RAAAVSEAEADFYBONSRYQLTLWGPE 652
Db 1811 SALDRDSRQHQHQLA-DSEAHRFDESGPLLRVCVMKAEREHYLVTLHHIV---TE 1866
Qy 553 GNILDYANKQLAGLVANVYTPRWRLFLLEALVDSVAQGIPIQQOHQFDKNVQLE 705
Db 1867 GWAMDIFARELGALYEAFLDOR-----ESPLELPVQ--YLDYSVWQRE 1908
```

RESULT 5

S18268  
delta-(L-alpha-aminoadipyl)-L-cysteinyI-D-valine synthetase - Streptomyces lactamdura  
C;Species: Streptomyces lactamdurans  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Nov-2000  
C;Accession: S18268; S15283; B38171  
R;Martin, J.F.  
submitted to the EMBL Data Library, January 1991  
A;Reference number: S18268  
A;Accession: S18268  
A;Molecule type: DNA  
A;Residues: 1-3649 <MAR>  
A;Cross-references: EMBL:X57310; NID:945005; PIDN:CAA40561.1; PID:945006  
R;Coque, J.J.R.; Martin, J.F.; Calzada, J.G.; Liras, P.  
Mol. Microbiol. 5, 1125-1133, 1991  
A;Title: The cephamycin biosynthetic genes pcbAB, encoding a large multidomain peptid  
genes in Acromonium chrysogenum and penicillium chrysogenum.  
A;Reference number: S15283; MUID:92065808  
A;Accession: S15283  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 224-940;1319-2010;2373-3307 <COQ>  
A;Cross-references: EMBL:X57310  
A;Note: the source is designated as Nocardia lactamdurans  
R;Coque, J.J.R.; Liras, P.; Laiz, L.; Martin, J.F.  
J. Bacteriol. 173, 6258-6264, 1991  
A;Title: A gene encoding lysine 6-aminotransferase, which forms the beta-lactam precu  
A;Reference number: A38171; MUID:92011390

A:Accession: B38171  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-23 <CO2>  
A:Cross-references: GB:S57006  
C:Genetics:  
A:Gene: pcbAB  
C:Superfamily: alpha-aminoacyl-pyl-cysteinyI-valine synthetase; acetate--CoA ligase homolog  
C:Keywords: carrier protein; cephamycin biosynthesis; phosphopantetheine; phosphoprotein  
F:298-758/Domain: acetate--CoA ligase homolog <ACLI>  
F:786-856/Domain: acyl carrier protein homolog <ACP1>  
F:1392-1844/Domain: acetate--CoA ligase homolog <ACLI2>  
F:1862-1932/Domain: acyl carrier protein homolog <ACP2>  
F:2446-2895/Domain: acetate--CoA ligase homolog <ACLI3>  
F:2912-2980/Domain: acyl carrier protein homolog <ACP>  
F:820,1896,2944/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 3.0%; Score 119; DB 1; Length 3649;  
Best Local Similarity 21.1%; Pred No. 16;  
Matches 174; Conservative 63; Mismatches 283; Indels 306; Gaps 37;

QY 16 AGAGGAAGDEAREAAVRALV-----ARLGGPAADFSVSVERALAAKPGLDYSLGG 69  
Db 1988 AALPGALGTLVRHRPALRTLLTKTDQGVRQVPIPADDDRLEVP----- 2031

QY 70 GGAARVRGSGTCVAAAAGLHRYLRDFCGCHVWGSQSLRPLRPPLPAVPGELTEATPNRY 129  
Db 2032 -----STVDSRAELDEVLTE-----RAGYVFLHEELPI-----RAEAFDHGD 2070

QY 130 RYONVCTQSYFVMMWDMARWERI-----DWMALNGINLAWSGOEALWQRYV 179  
Db 2071 EYLSVVVHSCDFGWSWDFRELAALLDGVPEADLGALRGT-----YGEFAVWQRY 2124

QY 180 LAIGLQAEINERFTGPFLWGRMNLHT-----WDGPLPPSHWIK-----QLYLQHRVLD 231  
Db 2125 -LT-GKRLAALTEFWT-----GALGGFETIALPLDHPRPFRFDYRGRELEFELDETTTE 2176

QY 232 QMSFGMT----- 239

Db 2177 ALRELARTARVLSYVLLGANCLMLNMYTQCHDLVVGTFPSANRGRPEEDRAVGFEEFANLLA 2236

QY 240 -----PVLPAFAGHPVAVTRVFPQVNTKMGWSGHFNCYSFLLAPEDPI-FPI 290  
Db 2237 LRVVDPAATLPAYRSVGEAV--VAAQVH-----GELPFQLYKELKVKDPSRHP 2287

QY 291 IGSFLRELKKEGTGHIYCADTFENOMPPSPSYLAATAATVAYEAMTAVDTEAVWLLQ 350  
Db 2288 LQLNFTLQNV-----SDH-----TSALTGYQDPS-----GGWTTTKFD-LSATMTETATGLA 2333

QY 351 GWL-----FQHQPFQWGPQAQTRAVLGAVPRGRLLLVLDLFAESQ-PVY 391  
Db 2334 GNLTYAASLFDDTSASGFIATFKHV-----LAEFASAAQTPIAQLTALDEPGQAALPDA 2388

QY 392 TRTASFOGQPIWCLNHFNGHGLFCALEAVNGGPEARLFPNSTVMGTG---MAPEGI 448  
Db 2389 TRRARRECGP-----GRCRRLF-----EEVAATWPDRAVAVHGDVRLTYREL 2430

QY 449 SQ--NEVVSYLMAELGNRKDP-----VPDLAAVWTSFAARRYGVSHRPGAGAA 493  
Db 2431 NERANRLAHLRSVAEPRADIELIALDLKSELTLIVATLWAKAGAAWMPIDPSYDDRIA 2490

QY 494 W-----RLLRSVYNCSEACRGNHNSPLVRRPSLQMNSTIWNRSVDVEAWRLLSTS 546  
Db 2491 FMLSDTGAKLVL-----AGEAHSRVRG-----LTSGDVLDLEQLDLTG 2529

QY 547 APSLATSPAFRYDLDLITRONAVQELVSLYVEARSAYLSKELASLLRAGCVLAYELLPAL 606  
Db 2530 EP--AENP-----VTETSTLAVAIY--TSGTTGPKPAVLVSHGVS----- 2567

QY 607 DEVLASDSRFLGSLWLEQAAAAVSEAEADFYEQNSRYQLTLWGPEGNIL----- 656  
Db 2568 DSFRAQLSGRYFGSPDESAEAVFL---ANVVFDSVEQLALSVLGGHKLVLPPPSAADD 2624

QY 657 ----DYANKOLAGLVANYTTPRWRLFLLEALVDSVAOQIPFQHQHFD 698  
Db 2625 PAFYELANRE--GL-----SVLSGTPTQVERFD 2650

RESULT 6  
O4HUD1  
debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 2D6 - human  
N:Alternate names: CYP2D6; cytochrome P450 isozyme 2D; cytochrome P450db1  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 03-Mar-2000  
C:Accession: S01199; A28883; J04156; A33629; A30335  
R:Gonzalez, F.J.; Skoda, R.C.; Kimura, S.; Umeno, M.; Zanger, U.M.; Nebert, D.W.; Gelboin, R.  
Nature 331, 442-446, 1988  
A:Title: Characterization of the common genetic defect in humans deficient in debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 2D6 - human  
A:Reference number: S01199; MUID:88122614  
A:Accession: S01199  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-497 <GON>  
A:Cross-references: EMBL:X08006; NID:g30450; PIDN:CAA30807.1; PID:g30451  
R:Gonzalez, F.J.; Vilbois, F.; Hardwick, J.P.; McBride, O.W.; Nebert, D.W.; Gelboin, R.  
Genomics 2, 174-179, 1988  
A:Title: Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid seq  
A:Reference number: A28883; MUID:88314109  
A:Accession: A28883  
A:Molecule type: mRNA  
A:Residues: 1-497 <GON>  
A:Cross-references: EMBL:M20403; NID:g181349; PIDN:AAA52153.1; PID:g181350  
R:Jiang, Q.; Voigt, J.M.; Colby, H.D.; Skoda, R.C.; Meyer, U.A.; Gonzalez, F.J.  
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995  
A:Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16)  
A:Reference number: J04153; MUID:95251703  
A:Accession: J04156  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-497 <JIA>  
R:Kimura, S.; Umeno, M.; Skoda, R.C.; Meyer, U.A.; Gonzalez, F.J.  
Am. J. Hum. Genet. 45, 889-904, 1989  
A:Title: The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification  
A:Reference number: A33629; MUID:90072069  
A:Accession: A33629  
A:Molecule type: DNA  
A:Residues: 1-373, 'V', 375-497 <KIM>  
A:Cross-references: EMBL:M33388; NID:g181303; PIDN:AAA53500.1; PID:g181304  
R:Manns, M.P.; Johnson, E.F.; Griffin, K.J.; Tan, E.M.; Sullivan, K.F.  
J. Clin. Invest. 83, 1066-1072, 1989  
A:Title: Major antigen of liver kidney microsomal autoantibodies in idiopathic autoimmune  
A:Reference number: A30335; MUID:89155788  
A:Accession: A30335  
A:Molecule type: mRNA  
A:Residues: 125-373, 'V', 375-485, 'T', 487-497 <MAN>  
A:Cross-references: EMBL:M24499; NID:g522194; PIDN:AAA36403.1; PID:g522195  
C:Genetics:  
A:Gene: GDB:CYP2D6  
A:Cross-references: GDB:132127; OMIM:124030  
A:Map position: 22q13.1-22q13.1  
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homolog  
C:Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; meta  
F:302-465/Domain: cytochrome P450 homolog <CYP>  
F:443/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 3.0%; Score 117.5; DB 1; Length 497;  
Best Local Similarity 21.7%; Pred. No. 1.2;  
Matches 97; Conservative 53; Mismatches 144; Indels 153; Gaps 22;

QY 345 AVVLL-----QGVLFQHQPFQWGPQAQTRAVLGAVPRGRLLLVLDLFAESQPVYTRTAS 396  
Db 14 AIFELLVLDLMHRRQRAARYPP---GFLPLPGL-----GNLLHVD----- 50

|  |     |  |                                |
|--|-----|--|--------------------------------|
| QY   | 397 | FOGQPFICWMLHFCGN-----HGLFGALEA-VANGGPEAAARL--FPNSTWV                   | 439                            |
|  |     | : :   :   :  |                                |
| Db   | 51  | FONTPYCFDQLRFRGDFVSLQLAWTVPVWVNLGLAAVREALVTHGEDTADPPVPITQIL            | 110                            |
|  |     | : :   :   :  |                                |
| QY   | 440 | GTFMAPEGI-----SONEVYYSIAELGWRKDPDPDLAAWTYSFAARYGVSHPD                  | 489                            |
|  |     | : :   :   :  |                                |
| Db   | 111 | GFGPSQGVFLARYCPAWRQRRSVSTLRNLGLGK---KSLQBWYTEAA-----                   | 158                            |
|  |     | : :   :   :  |                                |
| QY   | 490 | AGAAWRLLRSVYNCSGEACRGHNRSLVRRPSLQMNSTIWNKSDVFEAWRLLTLSAPS              | 549                            |
|  |     | : :   :   :  |                                |
| Db   | 159 | -----CLCAAFANHSGRPF--RPNGLLDKAV-----SNVI-----AS                        | 188                            |
|  |     | : :   :   :  |                                |
| QY   | 550 | LATSPAFRYD-----LLDLTRQAVQLSVLYEARSAYLSKELASLLRAGGVLAYE--               | 601                            |
|  |     | : :   :   :  |                                |
| Db   | 189 | LTCGRRRFYDDPRFLRLDLAQELKEESGFLREVLNAVPLVLIHPAL--AGKVLRFQKA             | 246                            |
|  |     | : :   :   :  |                                |
| QY   | 602 | LIPALDEVL-----ASDSRFLGSLWLEQARAAAVSEAEADFQEQNSRQLTLWGPE                | 652                            |
|  |     | : :   :   :  |                                |
| Db   | 247 | FLTQDELLETHRWTPPAQPPDRDLTFAEMEKAK-GNPSSFNDENLR-----                    | 296                            |
|  |     | : :   :   :  |                                |
| QY   | 653 | GNILDYANKQLAGLVANYTPTRWRLFLEALVDSVAQGIQFQHQFQKNVFQLEQ-----             | 706                            |
|  |     | : :   :   :  |                                |
| Db   | 297 | ---IIVADLFSAGMVTSTTLAWGLLMLHPDVQRRV---QQEIDDVIGQVRRPEMGDO              | 350                            |
|  |     | : :   :   :  |                                |
| QY   | 707 | -----AFVLSKORYPSQPRGDTVDL  | 726                            |
|  |     | : :   :   :  |                                |
| Db   | 351 | AHMPYTTAVTHEVQRF-----GDIVPL  | 372                            |
|  |     | : :   :   :  |                                |
| RESULT   | 7   |  |                                |
| D72593   |     | hypothetical protein APE1213 - Aeropyrum pernix (strain K1)            |                                |
| C:Species  |     | Aeropyrum pernix   |                                |
| C:Date   |     | 20-Aug-1999  | #sequence_revision 20-Aug-1999 |
| C:Accession  |     | D72593   | #text_change 20-Aug-1999       |
| R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Ya- |     |  |                                |
| DNA Res.   | 6   | 83-101, 1999   |                                |
| A:Title  |     | Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon |                                |
| A:Reference number   |     | A72450; MUID:99310339  |                                |
| A:Accession  |     | D72593   |                                |
| A:Status   |     | preliminary  |                                |
| A:Molecule type  |     | DNA  |                                |
| A:Residues   |     | 1-1374 <KAW>   |                                |
| A:Cross-references   |     | DDJ:AP0000061; NID:g5104821; PIDN:BAA80202.1; PID:g1043                |                                |
| A:Experimental source  |     | strain K1  |                                |
| C:Genetics   |     |  |                                |
| A:Gene   |     | APE1213  |                                |

```

Query Match      3.0%; Score 117; DB 2; Length 1374;
Best Local Similarity 20.0%; Pred. No. 5.6;
Matches 163; Conservative 78; Mismatches 249; Indels 324; Gaps 37;

QY      1 MEAVAVAAAVCVLLLAGAGG-----AGGDARARAAAVRA-----LVAR 38
      : : : : | : | | | | : | | | | |
Db      620 IDSLTLETGPHFTAAAGNGSGLGTVTAPATARALAVAAATDMAYLSLIQGYLP LLAG 679
      : : : : | : | | | | : | | | | |

QY      39 LLGPGPAADFSSV-SVERALAAKPLDITYSLGGGCAARVVRGSGTGVAACAGLHRYLRDFC 97
      : : : : | : | | | | : | | | | |
Db      680 LGGYGDPAYFSARGPSHAGAPKPGI-----AATGGFAYTTGTS 717
      : : : : | : | | | | : | | | | |

QY      98 GCHVAVSGSQLRPLRPLFAPVPGELTEATPNRYRYQNVCTQSYSFVWMDWARWERIDWM 157
      : : : | : | : | : | : | : | : |
Db      718 LDH--YTGGRLD--PRAAPLLFGGTSMATP-----M 744
      : : : | : | : | : | : | : | : |

QY      158 ALNCINIAL----AWSQCEAI----WDRVYLALGLTOAEINEFTGPAFLAWGRMNLHT 209
      | : : | : | : | : | : | : | : |
Db      745 AAGAAALAIQALKESLGERLGEUWRYVTALSMTAQ-----WRG--LPWGE MGN-- 793
      : : : | : | : | : | : | : | : |

QY      210 WDGPPLPPSWIHK-----QLVYLQHRVLQDMR-----SFGM-----TPV 241
      : : : | : | : | : | : | : | : |
Db      794 --GIVDAAGAIRLLTGVDQGVLYISATILSEAAQAGVAAAPGYPALLVWAGSGVETPY 851
      : : : | : | : | : | : | : | : |

```

|   |      |   |      |
|---|------|---|------|
| QY  | 242  | LFAFAGHVPEAVTRVFQVNVTKMGSHGHCNSYCSFLAPEDPIPIIGSLFLRELIX             | 301  |
| Db  | 852  | DIVLEGEQPVALKAVEPRLEVTVRSS-----VEIDLSSPVG                           | 887  |
| QY  | 302  | EGSTDHIIYGATTFNEMQPPSEPSYLAATATVYEAWTAVDTAEAVWLQGLWFOHQPOFW         | 361  |
| Db  | 888  | ESGYTYMTNAASIDPSMLPOGP--VEVSLTLPYE-----VFDRQG--RERTSTW              | 932  |
| QY  | 362  | GPAQIRAVL-----CAVPRGRLLVLDAFESQPVYTRTASFQOQPIWCMHLNF                | 410  |
| Db  | 933  | WEGYVYGVLLALLYWADLDGDLGDGDEFFILSIDKKSNVF-----RVF-----               | 976  |
| QY  | 411  | GNHGHLFGALFVANGVGPEARLFPNSTWGTGMAGPISQNEVWYSLMAELGW--RKDP           | 467  |
| Db  | 977  | -----LSNAQELAGAREA-----LQWS-EGVEEVVPLALAGLSWAGGTTRA                 | 1019 |
| QY  | 468  | VPDLAA--WVTSFAARYRGVSHDPDAGANRLLLRSLVSYNGSGEAC-----                 | 509  |
| Db  | 1020 | TVEIATISWESSA-----LSPATVEVEGEAAVAVNPPEGVVTGF                        | 1063 |
| QY  | 510  | -----RGHNSPLV---RRPSL-----QWNTSIWTVNRSDVF                           | 537  |
| Db  | 1064 | IVASTGVGEYRMPYTIILSPRYRPTGTVGHVYQGLGPDXYEEAAWVRGAFDYSWVYEDGD        | 1121 |
| QY  | 538  | EAWRLLTAPSLSATSPAFRYDLDLTROAVQELVSLVYEEARSAYLSKELASLLRAGGV          | 597  |
| Db  | 1122 | --WRIVPLELPGCLAVARLHWVEDEGRAYATNIDAY-----DYLNRPSITVSEAGEV           | 1173 |
| QY  | 598  | L--AYELLPALDEVLASDRF---LLGSWLE-----QARAAAVSEAEADFYEQNSRYQL          | 646  |
| Db  | 1174 | VVGVEENALASREASAANRFDDTLTGFWDQPCSGPGGETVLAAYNHA-----                | 1220 |
| QY  | 647  | TLWGPENGLDYANKQLAG-----LVANYTTP                                     | 673  |
| Db  | 1221 | ----PGRVLLVYRSVOYSGETARQPVVITLYHTP                                  | 1250 |
| RESULT  | 8    |   |      |
| T35238  |      | probable secreted cellulase - Streptomyces coelicolor               |      |
| C:Species:  |      | Streptomyces coelicolor   |      |
| C:Date:   |      | 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999 |      |
| C:Accession:  |      | T35238  |      |
| R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. |      |   |      |
|   |      | submitted to the EMBL Data Library, September 1998                  |      |
| A:Reference number:   |      | Z21572  |      |
| A:Accession:  |      | T35238  |      |
| A:Status:   |      | preliminary; translated from GB/EMBL/DBJ                            |      |
| A:Molecule type:  |      | DNA   |      |
| A:Residues:   |      | 1-973 <SEE>   |      |
| A:Cross-references:   |      | EMBL:AL031515; PIDN:CAA20643.1; GSPDB:GN00070; SCOEDB:SC            |      |
| A:Experimental source:  |      | strain A3(2)  |      |
| A:Gene:   |      | SCOEDB:SC5C7.31c  |      |

[illegible]

```
QY 137 TQSYFVWHDWARWEREIDWALNGINLALANSQGEAIWQRYLALGLTQAEINEFFTPG 196
Db 389 SEAYSILLW-----LQAMYGK-----TGD 408
QY 197 AFLAWGRGNLHTWDGPPSHIKQLYLQHRVLDQMSFGMTVPVLPFAFGHVPEAVT-R 255
Db 409 -----WSKFN-----ANEIMETWYIPTHADOPTNSSYNASKP--ATYAPELDPN 452
QY 256 VFPQVNVTKMSGWGHFNCSYSCSFLLAPEDPIFPPIIGSLFLRELKFKGTDHIIYA----- 311
Db 453 EXP-----APLDGTVSYGSDFIAGELKASAYGTDVYGMHWLQ 489
QY 312 -----EMOPSSSEPSYL-----AAATAYEAMTAVDTEA-----VM 347
Db 490 DVNDNTGYGNSCKEAGSDTGPYINTFORGAESVWETVPQTCDAFKYGGKNGYLD 549
QY 348 LLOG-----WLFQHQFQFWGPAQIRAVLGAVPRGRLLVLDLFAESQ-----PVYTRT 394
Db 550 LFTGDASYAKQKFTNAPD-----ADARAVQAA-----YWDIWAAGEQCKSDEISATLDK 599
QY 395 ASFOGQPFITWCLMHNFGNGHGLFGALEAVNGGPEAAFLPNTWMTGTMAGEGISONEVV 454
Db 600 AAKMGDYLYAMFDKYFKKV-----NCVGPSACPAGTGKDSH 638
QY 455 YSLMAELGWKRKDPVPDLAAWTSFAARRYGVSHPDAGAAWRLLLRSVYVNCSGEACRGHNR 514
Db 639 Y-----LLSWYANG-----GAVDTSAGWAWRI-----GSSHTHGGYQ 671
QY 515 SPLVR-----RPSLQWNTSIWYNRSD-VFEAWRLLLTS-----APSLAT 552
Db 672 NPLAAAYALSTDADLKPKSATGSDWAKSLDRQVEFYRWLQSDGEGATAGATNSWAGRYAT 731
QY 553 SPAFRYDLDLTRQAOVELSVLYEB-----ARSAYLSKELASLLRAGGVLA 599
Db 732 PPA-----GPTFYGHYDEKPYHDPSPNQWFGFQAWSERVEYVYQSG--- 777
QY 600 YELLPALDEVLSRFLGSLWLEQARAASAEADFEQNSRYOLTLL-----W--- 649
Db 778 -----DAGAKAVLDKWDWALSETVNPQDTE-----RIPSTLQWSQPDWNA 821
QY 650 ---GPEGN---ILDYANKQLAGLVA-----NYTTPR-----WRLFLEALVDV--- 686
Db 822 SSPGANGDLHVEADYDND--VGVAAAYAKTLTYADRSGDTEAATAKALDGMWENNQ 879
QY 687 -AOGIPEQHOHQFNKVFQLEQAFVLSKQRYPSQPRGDTVDLAK-----KIFLYYPCW 738
Db 880 DALGIAPETRADYNRF--DDGIYVSGWSGTMNGDVTVDASSTFASIRSFYQDDPAW 935

RESULT 9
ComEC/Rec2 family protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: C87485
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87485
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <STO>
A:Cross-references: GB:AE005673; NID:g13423355; PIDN:AAK23879.1; GSPDB:GN00148
C:Genetics:
C:Gene: CCI904

Query Match 2.9%; Score 115; DB 2; Length 725;
Best Local Similarity 21.9%; Pred. No. 3.2;
Matches 137; Conservative 67; Mismatches 253; Indels 170; Gaps 29;
```

```
QY 1 MEAVAVAAAVGVLLAGAGGAGDEAR--EAAAVRALVARLLGPGPAADFVSYSVERALAA 58
Db 54 LEVRVEPALWGLTLLAGVIGLAANGARRWNAPPLVLLVLLGLIAGGAAGAAKVRSERVA 113
QY 59 KPGL-----DYSLGGG-----AARVRVSGTGVAAAAGLLHRYLRDFCGCH 100
Db 114 APVMDGERIVRRVDGVVVDVSPGAGGORLLIAPRVSG-----LSPEDTPRRIIVTIGEN 169
QY 101 -VANGSOLRL-----PPLPVAVPGLTEATPNRYRYQNVCTQSYSF---VWMDWARWE 151
Db 170 PVKPGQAGIGURAMLGPPPPAAG-----AYDFARDAWED----- 205
QY 152 REIDWMLNGINLALANSQGEAIWQRYLALGLTQAEINEFFTPVLPFAFGHGRMNHLTWD 211
Db 206 -----SIGGVGFAIGDIGEVTLDQPP-LRLRLVMA-VNAFRWDLAQRLLARMGPTSGGI 257
QY 212 GPLPPSHIKQLYLQHRVLDQMSFGMTVPVLPFAFGHVPEAVTRVFPQVNVTKMSGWCHF 271
Db 258 GAAMVTGH--EAWISETQTEAMRASGLAHILISISGLHM--AIVGGFVTV-VVRMGV----- 308
QY 272 NCYSYCSFLLAPEDPIFPPIIGSLFLRELKFKGTDHIIYGADTFNEMQPPSEPSYLA-- 329
Db 309 -AAWPWLALRAPGKKIAASAG-----LVSVLGYLIVSGA-----PPPAERAAITASV 354
QY 330 ATTAYVEAMTAVDTEAVWMLQGWLFQHQFQFWGPAQIRAVLGAVPRGRLLVLDLFAESOP 389
Db 355 AFLAILFORRAITLHGLAALAILLKPETAGEPGFQMSFAAT-----AALVALAESWP 409
QY 390 VYTRTAS---FOGQPFIT--WCMLH-----NFGNGHGLFGALEAVNGGPEAAR 431
Db 410 KPVRELSTFWNIQGGQATATWLVASIAASLVAGLATATAFAMQH--FNRV-AWVGLPANLA 466
QY 432 LPPNSTWV-----GTGMAPEGISQNEVYVSLMAELGWKRKDPVPDLAAWTSFPAARYG 484
Db 467 VSPSSSFVIMPELAIGTVLEPGLSA-----PFLAVAGWIDVMLNVA---GLFSEAHG. 517
QY 485 VSHPDAGA-----AWRLLLRSV-----YNCSG 506
Db 518 AQHIVASAPPOVLLVAFGLMVLCWRLWIGAPLALAVALNPRPTPPDIAADGAT 577
QY 507 EACRGHNRSPVLRPRPSLOMNTSIWYNR 533
Db 578 AAVRSCNAALLRTDAKREGAELMARR 604

RESULT 10
T31293
aldehyde dehydrogenase homolog - Sphingomonas aromaticivorans plasmid pNL1
C:Species: Sphingomonas aromaticivorans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T31293
R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.
submitted to the EMBL Data Library, July 1998
A:Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas arom
A:Reference number: Z20992
A:Accession: T31293
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-479 <ROM>
A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378434; PIDN:AAD04017.1
C:Genetics:
C:Gene: plasmid pNL1
C:Note: nahF
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 2.9%; Score 114; DB 2; Length 479;
Best Local Similarity 23.3%; Pred. No. 2.1;
Matches 112; Conservative 52; Mismatches 165; Indels 152; Gaps 24;
```

```
QY 222 QLYLQHRVLDQMSFGMTVPVLPFAFGHVPEAVTRVFPQVNVTKMGS---WGHFNCSYSCS 278
Db 222 QLYLQHRVLDQMSFGMTVPVLPFAFGHVPEAVTRVFPQVNVTKMGS---WGHFNCSYSCS 278
```



|    |     |   |     |
|----|-----|---|-----|
| QY | 349 | LOGWLFQHQPOFWGPAQIRAVLGAVP-----RGRLLVLDLFAESQPVYTRTA---SFQG | 399 |
| Db | 40  | LAGWAREAGVELRG---LRLGICEAPGEWLDGGNLLILDTPTDRQAVEELGERLQG    | 96  |
| QY | 400 | --OPTIWCLINFGNGHGLFGALEA-----VNGSP-----EAARLPFNSTMVG        | 440 |
| Db | 97  | GTQPMI-----RVGGGGPGGNGLPALGGRLGVYVANGGEANLRLEFVRRWHAGLTV    | 151 |
| QY | 441 | TCMAPEGISQ-----NEVVYSIMAE-LGWKDPDPLAAWVT-----SFAARRYGVSH    | 487 |
| Db | 152 | ALPAPQPLQAQGFYHPDAPPAFVLADYLAW-----GASRWASDAPRIAFILPRAIAD   | 205 |

| Query Match           | 2.8%;                | Score 112;                       | DB 2;                             | Length 1323;      |
|-----------------------|----------------------|----------------------------------|-----------------------------------|-------------------|
| Best Local Similarity | 22.9%;               | Pred. No. 12;                    |                                   |                   |
| Matches 106;          | Conservative .48;    | Mismatches 157;                  | Indels 152;                       | Gaps              |
| QY 162                | INLAWSGOEAIQWRVYLALG | LTQAEINEFFTGPAFLAWGMGNLH         | -----TWDGPLP                      | 215               |
| Db                    | : : : : :   :   :    | : : : : :   :   :                | : : : : :   :   :                 | : : : : :   :   : |
| QY 48                 | LNVLVFSFG            | -----AYAAEAARL                   | -----GPAVAAVRSPLDVRPVALVINGSDP    | 93                |
| Db                    | : : : : :   :   :    | : : : : :   :   :                | : : : : :   :   :                 | : : : : :   :   : |
| QY 216                | PSWHIKOLVYQ          | -----HRLV                        | -DOMRSFGTTPVLPAFAGHVPAVTRVFPQNVNT | 263               |
| Db                    | : : : : :   :   :    | : : : : :   :   :                | : : : : :   :   :                 | : : : : :   :   : |
| QY 94                 | PS                   | -----LVQLCDLLSLGRVHG             | VFEDDSRAPAVAPILDFLSAQTSLPIVAVHG   | 143               |
| Db                    | : : : : :   :   :    | : : : : :   :   :                | : : : : :   :   :                 | : : : : :   :   : |
| QY 264                | KMGSGWHFNCSCSFLLAPED | PIPTGSLPRLIKEFGTQDHIYGADTFNEMQ   | -----                             | 318               |
| Db                    | : : : : :   :   :    | : : : : :   :   :                | : : : : :   :   :                 | : : : : :   :   : |
| QY 144                | -----GAALVLP         | PKPK-----GSTFLQ                  | -----LGSSTEQQLQVIFEV              | 175               |
| Db                    | : : : : :   :   :    | : : : : :   :   :                | : : : : :   :   :                 | : : : : :   :   : |
| QY 319                | -PPSSPSYLAATA        | -----VYEAMTAVDTEAVWLLOGNLWFQHQPF | -WGPQAQIRAVLGA                    | 372               |
| Db                    | : : : : :   :   :    | : : : : :   :   :                | : : : : :   :   :                 | : : : : :   :   : |

Db 176 LBEYDWTSEVATVTRAPCHRAFLSVIEVLTG--SLVGW--EHRGALTDPGAEAVLGA 231  
Qy 373 VPRG-----RLVLDFAESOPVY--TRTASFQGGPFTWML-HNFGNHLGFGALEAVN 424  
Db 232 QLRYSVAQIRLLFC-AREAEVPFAEAEAGLTGPGYWMFVGPQLAGGG-----S 282  
Qy 425 GQPEARLPNSTWGTGMAPGISQNEVYSLMAELGWRKDPVPLAAWNTSPA----- 479  
Db 283 GYPGEPLLPGGSPPLAGL-----FAVASAGRODLARVAVGAVVARGAQA 330  
Qy 480 -ARRYGVSHPDAGAAWRLRLRSVNCSEACRGHNRSLVRPSLQ--MNTSIWYNSRD 535  
Db 331 LLRDVGLF-PEL-----GHDCRTQNRTH--RGESLHRYFMNIT-WDNRDY 371  
Qy 536 VFEAWRLLTSPASLATSAPFYDLDTLQRAVQELVSLYEE 578  
Db 372 SFNEDGFLVN--PSLV-----VISLTRDRTWVGVGSWEQO 404

RESULT 13  
H87316  
hypothetical protein CC0545 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: H87316  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: H87316  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1027 <STO>  
A:Cross-references: GB:AE005673; NID:g13421736; PIDN:AAK22532.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0545

Query Match 2.8%; Score 111; DB 2; Length 1027;  
Best Local Similarity 22.5%; Pred. No. 10;  
Matches 157; Conservative 47; Mismatches 217; Indels 278; Gaps 38;

Qy 135 VCTOSYSFVN-----WDWAWERE-----IDWMLNGINLAL-----AWSQGEATW 175  
Db 68 VCDVSATGQWLLRASERAWDLKRLNDEGSLRKAIGWRRLDGADPATADLLDALLGKGAFS 127  
Qy 176 QRVYALGLTQAEINEFTTGAFLWGRMGNLHTWDGPIPPSWH-IKQLYLQHRVLDQMR 234  
Db 128 PKA-MARMLERIEAQEAAGPEL---RRLATLEWSGALAPAAALRLAQALNRRTDOAR 183  
Qy 235 SFGMTPLPAPAGHVPEAVTRVPPVNVTKMSGHNCYSYCSFLLAP-EDPI----- 287  
Db 184 R-GELLADGFGQR-----EQAVEILTGW-----IRNPODDPIRVRLY 221  
Qy 288 ---FPPIIG-SLFLRELKIFGTGDI-----YGATFNEM-----QPSSEP 324  
Db 222 VSLGPAIGKSTLLEHLVRQAQGVETVWVRLDFDRASLDVLDQLGLSVEVARQVASQLP 281  
Qy 325 SYLAATATVYEAAT-----AVDTEAVWL 348  
Db 282 SAAADLRQARLEAITTTQETRLKGGSGRGTPIALIQALGVLQAGSAGRLRVWVLDLTLEVL 341  
Qy 349 LOGWLFQHQPOFWGPAQITRAVL-----GAVPRGRLL--VLDFLFAES----- 387  
Db 342 GRGETHPQLDFDLQDLARVLVPLVVVACGRGRALEPVDRAETLNLDRLTDAEAERL 401  
Qy 388 -----QPVYTRT-ASFQGGPFTWMLHNFNGHNG-----LFGALEAVNGG 426  
Db 402 LSAADVPPALMTQVIALAEGDPLQLRLATLARDEQAGLTGKGATITLYRSLEARLGG 461  
Qy 427 PEA--ARLFP-----NSTMVTGMAP-----EGISQNEVYVSLMAEL-----GW 463

Db 462 RHAKLARFIPLFRFENSELLGAVNPVVLGERLSPGAAA-ALMAELKDOFWLIAGTEGW 520  
Qy 464 RKDPVPD-----LAAWVTSFAARRYGVSHPDAGAAWRLRLRSVNCSEACRGHNRSPVLR 519  
Db 521 -LTPQDRLRLRACKLYDEHARRRTGALHRRR-ARK-----FAE 556  
Qy 520 RPSLQMTNTSIWYNSRVFEAWRLLTSA--PSLATSAPFR-YDLLDTLQRAVQELVSLY 575  
Db 557 RPELWAGAESLYHR--LOATRWAGPDALKGMNLAAGVAFQASADLEELP----- 602  
Qy 576 YEEARSAYLSKELASLRRAGGVLAYE-----LLP-----ALDEV-----LASDSR 615  
Db 603 -EAARDA-----LLRARGERSYEGRGDALTPSGEVSPPAAADELRMLTSGKGLAEASH 653  
Qy 616 FL-----LGSWLEQARAAA 629  
Db 654 FYGRVFRPAALDPTSPADVALTLTLWRTGRTWEARRLAA 692

RESULT 14  
F87285  
conserved hypothetical protein CC0295 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: F87285  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: F87285  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-285 <STO>  
A:Cross-references: GB:AE005673; NID:g13421436; PIDN:AAK22282.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0295

Query Match 2.8%; Score 109.5; DB 2; Length 285;  
Best Local Similarity 23.7%; Pred. No. 2.1;  
Matches 80; Conservative 36; Mismatches 90; Indels 131; Gaps 17;

Qy 328 AAATTAYEAMTAVDTAEAVMLQGLFQIQPFQWGPQAIQRAVLGAVPGRLLVLDLFAES 387  
Db 3 AAPTAAWAGLTAYSWLATTICAAWF-----ARRFALADGRVLSE----- 43  
Qy 388 QPVYTRTASFQGGP-----IWMCLHNFNGNHLFGALEA----- 422  
Db 44 -----OSLAWQGAITYAANLPVIGIVWLVLRRFGA--GLRGVLAAPATGLVIVPLEALVAS 96  
Qy 423 -----VNGSGPAAR-----LFPNSTWVTGTM-----APEGISQNEVYVSLMAE- 460  
Db 97 LIDOTEFAGGDLAERALGRAPVCILLHTAIVAGVLAAGAAHRRRAAEASARNTLLQAAALAA 156  
Qy 461 -----LGRWKDPVPDLAAWNTSFAARRYGVSHPDAGAAWRLRLRSVNCSS 505  
Db 157 RATSVLDVAERLMMVAGARRVPV-DITAVENFGAANDYVVVHWDGREG---LMRATLQ-S 211  
Qy 506 GEA-----CRGHNRSPVLRPSLQMTNTSIWYNSRVDFEA-----WRLLTSPASLA 551  
Db 212 LEARLDPLRFARAH-RSALV-----NLSKVSQAQPLSDGNSRLTLASGAEVV 257  
Qy 552 TSPAFRYDLTD-ITROAVQELVSLYEEARSAYLSKE 587  
Db 258 TSRTYRDALLKRLGROA-----SDSOASMSSE 285

RESULT 15  
T47641  
hypothetical protein T15C9.20 - Arabidopsis thaliana

Search completed: September 21, 2002, 12:24:47  
Job time: 62 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2002, 12:24:50 ; Search time 13.28 seconds  
(without alignments)  
2166.312 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939

Sequence: 1 MEAVAVAAAGVLLLAGAGG.....VDLAKIFLYKYPGWAGSV 743

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID        | Description         |
|------------|-------|-------------|--------|--------------|---------------------|
| 1          | 3939  | 100.0       | 743    | 1 ANAG_HUMAN | P54802 homo sapien  |
| 2          | 123.5 | 3.1         | 497    | 1 CPDH_MACFA | Q29488 macaca fasc  |
| 3          | 119   | 3.0         | 3649   | 1 ACVS_NOCIA | P27743 nocardia la  |
| 4          | 117.5 | 3.0         | 497    | 1 CPD6_HUMAN | P10635 homo sapien  |
| 5          | 110.5 | 2.8         | 1029   | 1 YK95_RHIME | Q52999 rhizobium m  |
| 6          | 109.5 | 2.8         | 1377   | 1 RHSA_ECOLI | P16916 escherichia  |
| 7          | 109.5 | 2.8         | 1411   | 1 RHSE_ECOLI | P16917 escherichia  |
| 8          | 108   | 2.7         | 726    | 1 CATA_SALTY | P17750 salmonella   |
| 9          | 104.5 | 2.7         | 726    | 1 CATA_ECOLI | P13029 escherichia  |
| 10         | 103   | 2.6         | 722    | 1 P85B_RAT   | Q63788 rattus norv  |
| 11         | 101   | 2.6         | 698    | 1 YMCA_ECOLI | P75882 escherichia  |
| 12         | 101   | 2.6         | 3519   | 1 OL56_STRAT | Q07017 streptomyce  |
| 13         | 100   | 2.5         | 851    | 1 ENV_HV2D1  | P17755 human immun  |
| 14         | 100   | 2.5         | 1323   | 1 NME4_MOUSE | Q03391 mus musculus |
| 15         | 100   | 2.5         | 1336   | 1 NME4_HUMAN | O15399 homo sapien  |
| 16         | 99.5  | 2.5         | 1092   | 1 LIFR_MOUSE | P42703 mus musculus |
| 17         | 99    | 2.5         | 624    | 1 SIR_SYNP7  | P30008 synechococc  |
| 18         | 99    | 2.5         | 1230   | 1 UGSA_SOLTU | Q43846 solanum tub  |
| 19         | 99    | 2.5         | 1587   | 1 LMG3_HUMAN | O9y6n6 homo sapien  |
| 20         | 98.5  | 2.5         | 512    | 1 NUOM_RHOCA | P22367 penicillium  |
| 21         | 98.5  | 2.5         | 1774   | 1 MSAS_PENPA | P18094 human immun  |
| 22         | 98    | 2.5         | 860    | 1 ENV_HV2BE  | O62645 rattus norv  |
| 23         | 98    | 2.5         | 1323   | 1 NME4_RAT   | P33144 emericella   |
| 24         | 98    | 2.5         | 2067   | 1 B1MB_EHANI | Q9kdw8 bacillus ha  |
| 25         | 97.5  | 2.5         | 497    | 1 GLPK_BACHD | P10231 herpes simp  |
| 26         | 97.5  | 2.5         | 693    | 1 UL47_HSV1  | P29958 actinoplane  |
| 27         | 97.5  | 2.5         | 786    | 1 AAC_ACTUT  | P27599 bovine herp  |
| 28         | 97.5  | 2.5         | 842    | 1 VGLH_HSVBC | P79621 mus musculus |
| 29         | 97.5  | 2.5         | 1078   | 1 C2TA_MOUSE | P72059 mycobacteri  |
| 30         | 97    | 2.5         | 1094   | 1 EMBG_MYCTU | P20806 drosophila   |
| 31         | 97    | 2.5         | 2594   | 1 7LES_DROVI | P05877 human immun  |
| 32         | 96.5  | 2.4         | 856    | 1 ENV_HV1MN  | P57716 mus musculus |
| 33         | 96    | 2.4         | 708    | 1 NICA_MOUSE |                     |

|    |      |     |      |              |                    |
|----|------|-----|------|--------------|--------------------|
| 34 | 96   | 2.4 | 2314 | 1 PTP2_HUMAN | P23471 homo sapien |
| 35 | 95.5 | 2.4 | 330  | 1 RCME_ERVSP | P26279 erythrobaet |
| 36 | 95.5 | 2.4 | 914  | 1 PERT_RAT   | P14650 rattus norv |
| 37 | 95   | 2.4 | 344  | 1 LIC1_BURCE | P22089 burkholderi |
| 38 | 94.5 | 2.4 | 664  | 1 UL47_HSV1F | P08313 herpes simp |
| 39 | 94.5 | 2.4 | 761  | 1 CTPA_MYCTU | Q06457 mycobacteri |
| 40 | 94.5 | 2.4 | 866  | 1 NASA_KLEPN | Q06457 klebsiella  |
| 41 | 94   | 2.4 | 500  | 1 CPDG_CAVPO | Q64403 cavia porce |
| 42 | 94   | 2.4 | 990  | 1 AMPM_MANSE | O11001 manduca sex |
| 43 | 94   | 2.4 | 1616 | 1 APXL_HUMAN | Q13796 homo sapien |
| 44 | 94   | 2.4 | 2303 | 1 POLG_TMEVB | P08544 t genome po |
| 45 | 93.5 | 2.4 | 276  | 1 SX21_HUMAN | Q9y651 homo sapien |

#### ALIGNMENTS

|            |  |
|------------|--|
| RESULT     | 1  |
| ANAG_HUMAN |  |
| ID         | ANAG_HUMAN   |
| AC         | P54802; STANDARD; PRT; 743 AA.   |
| DT         | 01-OCT-1996 (Rel. 34, Created)   |
| DT         | 01-OCT-1996 (Rel. 34, Last sequence update)  |
| DT         | 16-OCT-2001 (Rel. 40, Last annotation update)  |
| DE         | Alpha-N-acetylglucosaminidase precursor (EC 3.2.1.50) (N-acetyl-alpha-glucosaminidase) (NAG).                                  |
| DE         | NAGLU OR UPHSDI.   |
| GN         | Homo sapiens (Human).  |
| OS         | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  |
| OC         | Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.  |
| OX         | NCBI_TaxID=9606;   |
| RN         | [1]  |
| RC         | SEQUENCE FROM N.A.   |
| RC         | TISSUE=Testis;   |
| RX         | MEDLINE=96234097; PubMed=8650226;  |
| RA         | Zhao H.G., Li H.H., Bach G., Schmidtchen A., Neufeld E.F.;   |
| RT         | "The molecular basis of Sanfilippo syndrome type B.";  |
| RL         | Proc. Natl. Acad. Sci. U.S.A. 93:6101-6105(1996).  |
| RN         | [2]  |
| RP         | SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  |
| RX         | MEDLINE=96372812; PubMed=8776591;  |
| RA         | Weber B., Blanch L., Clements P.R., Scott H.S., Hopwood J.J.;  |
| RT         | "Cloning and expression of the gene involved in Sanfilippo B syndrome (mucopolysaccharidosis III B).";                         |
| RL         | Hum. Mol. Genet. 5:771-777(1996).  |
| RN         | [3]  |
| RP         | SEQUENCE FROM N.A.   |
| RA         | Zhao Z., Yazdani A., Shen Y., Sun Z.S., Bailey J., Caskey C.T.,  |
| RL         | Lee C.C.;  |
| RN         | Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  |
| RN         | [4]  |
| RP         | VARIANTS MPS-IIIB H-92; S-115; C-140; K-153; L-358; V-664 AND R-682.   |
| RX         | MEDLINE=98107938; PubMed=9443878;  |
| RA         | Schmidtchen A., Greenberg D., Zhao H.G., Li H.H., Huang Y., Tieu P.,   |
| RA         | Zhao H.-Z., Cheng S., Zhao Z., Whitley C.B., di Natale P.,   |
| RA         | Neufeld E.F.;  |
| RT         | "NAGLU mutations underlying Sanfilippo syndrome type B.";  |
| RL         | Am. J. Hum. Genet. 62:64-69(1998).   |
| RN         | [5]  |
| RP         | VARIANTS MPS-IIIB.   |
| RX         | MEDLINE=99133861; PubMed=9950362;  |
| RA         | Bunge S., Knigge A., Steglich C., Kleijer W.J., van Diggelen O.P.,   |
| RA         | Beck M., Gal A.;   |
| RT         | "Mucopolysaccharidosis type IIIB (Sanfilippo B): identification of 18 novel alpha-N-acetylglucosaminidase gene mutations.";    |
| RL         | J. Med. Genet. 36:28-31(1999).   |
| CC         | !- FUNCTION: INVOLVED IN THE DEGRADATION OF HEPARAN SULFATE.   |
| CC         | !- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-acetyl-D-glucosamine residues in N-acetyl-alpha-D-glucosaminides. |
| CC         | !- SUBUNIT: MONOMER AND HOMODIMER.   |
| CC         | !- TISSUE SPECIFICITY: LIVER, OVARY, PERIPHERAL BLOOD LEUKOCYTES,  |
| CC         | !- TESTIS, PROSTATE, SPLEEN, COLON, LUNG, PLACENTA AND KIDNEY.   |
| CC         | !- DISEASE: DEFECTS IN NAGLU ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS  |

TYPE IIIB (MPS-IIIB) (ALSO KNOWN AS SANFILIPPO B SYNDROME), AN AUTOSOMAL RECESSIVE DISORDER WHOSE CLINICAL FEATURES ARE SEVERE MENTAL DETERIORATION BUT MILD SOMATIC MANIFESTATIONS IN CHILDHOOD, AND DEATH IN THE SECOND DECADE. BIOCHEMICALLY, THIS DISEASE IS CHARACTERIZED BY UNDEGRADED OR PARTIALLY DEGRADED HEPARAN SULFATE WHICH ACCUMULATES IN LYOSOMES AND IS EXCRETED IN URINE.

-----

THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; U43572; AAC50512.1; -;  
CC EMBL; U43573; AAC50513.1; -;  
CC EMBL; U40846; AAB06188.1; -;  
CC EMBL; L78464; AAB36604.1; -;  
DR MIM; 252920; -;  
KW Hydrolase; Glycosidase; Glycoprotein; Mucopolysaccharidosis;  
KW Signal; Disease mutation; Polymorphism.  
FT SIGNAL 1 23  
FT CHAIN 24 743  
FT CHAIN 59 743  
FT DOMAIN 68 71  
FT DOMAIN 84 87  
FT CARBOHYD 134 134  
FT CARBOHYD 261 261  
FT CARBOHYD 272 272  
FT CARBOHYD 435 435  
FT CARBOHYD 503 503  
FT CARBOHYD 526 526  
FT CARBOHYD 532 532  
FT CARBOHYD 79 79  
FT VARIANT 92 92  
FT VARIANT 100 100  
FT VARIANT 115 115  
FT VARIANT 140 140  
FT VARIANT 142 142  
FT VARIANT 153 153  
FT VARIANT 243 243  
FT VARIANT 277 277  
FT VARIANT 280 280  
FT VARIANT 292 292  
FT VARIANT 358 358  
FT VARIANT 452 452  
FT VARIANT 482 482  
FT VARIANT 561 561  
FT VARIANT 565 565  
FT VARIANT 643 643  
FT VARIANT 664 664  
FT VARIANT 674 674

POLY-GLY.  
POLY-ALA.  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
G -> C (IN MPS-IIIB).  
/FTID=VAR\_008979.  
Y -> H (IN MPS-IIIB).  
/FTID=VAR\_005007.  
H -> R.  
P -> S (IN MPS-IIIB).  
/FTID=VAR\_005008.  
Y -> C (IN MPS-IIIB).  
/FTID=VAR\_005009.  
MISSING (IN MPS-IIIB).  
/FTID=VAR\_008981.  
E -> K (IN MPS-IIIB).  
/FTID=VAR\_005010.  
P -> L (IN MPS-IIIB).  
/FTID=VAR\_008982.  
C -> F (IN MPS-IIIB).  
/FTID=VAR\_008983.  
L -> P (IN MPS-IIIB).  
/FTID=VAR\_008984.  
G -> R (IN MPS-IIIB).  
/FTID=VAR\_008985.  
P -> L (IN MPS-IIIB).  
/FTID=VAR\_005011.  
E -> K (IN MPS-IIIB).  
/FTID=VAR\_008986.  
R -> W (IN MPS-IIIB).  
/FTID=VAR\_008987.  
L -> R (IN MPS-IIIB).  
/FTID=VAR\_008988.  
R -> Q (IN MPS-IIIB).  
/FTID=VAR\_008989.  
R -> H (IN MPS-IIIB).  
/FTID=VAR\_005012.  
A -> V (IN MPS-IIIB).  
/FTID=VAR\_005013.  
R -> H (IN MPS-IIIB).

FT VARIANT 682 682 /FTID=VAR\_005014.  
FT L -> R (IN MPS-IIIB).  
FT /FTID=VAR\_005015.  
FT E -> K (IN MPS-IIIB).  
FT /FTID=VAR\_008990.  
FT G -> R.  
FT /FTID=VAR\_008991.  
FT CONFLICT 551 551 A -> L (IN REF. 2; AA SEQUENCE).  
FT CONFLICT 553 553 S -> L (IN REF. 2; AA SEQUENCE).  
SQ SEQUENCE 743 AA; 82166 MW; 6D8D6A42C7BA6083 CRC64;  
  
Query Match 100.0%; Score 3939; DB 1; Length 743;  
Best Local Similarity 100.0%; Pred. No. 3.4e-279; Indels 0; Gaps 0;  
Matches 743; Conservative 0; Mismatches 0;  
  
QY 1 MEAVAVAAVGVLLLAGAGGAAGDEAREAAARVALVARLLGPGPAADFSVSVERALAAKP 60  
DB 1 MEAVAVAAVGVLLLAGAGGAAGDEAREAAARVALVARLLGPGPAADFSVSVERALAAKP 60  
QY 61 GLDYSYSLGGGGAARVRVRSSTGVAAAAGLHRYLRDFCGCHVAVWSGSQLRPLRPVAPVGE 120  
DB 61 GLDYSYSLGGGGAARVRVRSSTGVAAAAGLHRYLRDFCGCHVAVWSGSQLRPLRPVAPVGE 120  
QY 121 LTEATPNRYRYQNVCTOSYSFVWMDWARWEREDWMALNGINLALAWSGOEAIWQRYVL 180  
DB 121 LTEATPNRYRYQNVCTOSYSFVWMDWARWEREDWMALNGINLALAWSGOEAIWQRYVL 180  
QY 181 ALGLTQAEINEFFFTGPAFLAWGRMGNLHTWDGPPSHHKKOLYLQHRVLDQMRSGMTP 240  
DB 181 ALGLTQAEINEFFFTGPAFLAWGRMGNLHTWDGPPSHHKKOLYLQHRVLDQMRSGMTP 240  
QY 241 VLPAPAGHVPEAVTRVFPQVNVTKMSWGHENCSYSCSFLAPEDPIPIIGSLFLREL 300  
DB 241 VLPAPAGHVPEAVTRVFPQVNVTKMSWGHENCSYSCSFLAPEDPIPIIGSLFLREL 300  
QY 301 KEFGTDHYGADTFNEMOPPSSEPSYLAATAATVAYEAMTAVDTEAVLWLLQGMFLFOHQPF 360  
DB 301 KEFGTDHYGADTFNEMOPPSSEPSYLAATAATVAYEAMTAVDTEAVLWLLQGMFLFOHQPF 360  
QY 361 WGAQIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFGQPFQWCMHNFNGHGLFGAL 420  
DB 361 WGAQIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFGQPFQWCMHNFNGHGLFGAL 420  
QY 421 EAVNGSPEAARLFPNSTVMGTGMAPEGISQNEVYSLMAELGWRKDPVDPDLAAWVTSFAA 480  
DB 421 EAVNGSPEAARLFPNSTVMGTGMAPEGISQNEVYSLMAELGWRKDPVDPDLAAWVTSFAA 480  
QY 481 RRYGVSHPDAGAAWRLRLRSVYNCSGEACRGNHRSPLVRRPSLQMNSTSIWYNRSDVFEAW 540  
DB 481 RRYGVSHPDAGAAWRLRLRSVYNCSGEACRGNHRSPLVRRPSLQMNSTSIWYNRSDVFEAW 540  
QY 541 RLLLTSPASLSPAPRYDLDDLTRQAVQELVSLYEEARSAYLSKELASLLRAGGVLAY 600  
DB 541 RLLLTSPASLSPAPRYDLDDLTRQAVQELVSLYEEARSAYLSKELASLLRAGGVLAY 600  
QY 601 ELLPALDEVLASDRFLGSLWLEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYAN 660  
DB 601 ELLPALDEVLASDRFLGSLWLEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYAN 660  
QY 661 KQLAGLVANYTTPRWLFLEALVDSVAQGIPIQOQHFQDKNFTQLEQAFVLSKORYPSQPR 720  
DB 661 KQLAGLVANYTTPRWLFLEALVDSVAQGIPIQOQHFQDKNFTQLEQAFVLSKORYPSQPR 720  
QY 721 GDTVDLAKKIFLKYYPGWAGSW 743  
DB 721 GDTVDLAKKIFLKYYPGWAGSW 743  
  
RESULT 2  
CPDH\_MACFA STANDARD; PRT; 497 AA.  
ID CDPDH\_MACFA  
AC Q29488;

01-NOV-1997 (Rel. 35, Created)  
 01-NOV-1997 (Rel. 35, Last sequence update)  
 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytochrome P450 2D17 (EC 1.14.14.1) (CYP2D17).  
 GN CYP2D17  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 CC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Lawton M.P., Laddison K.J., Speirs A.A., Mankowski D.C., Tweedie D.J.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
 CC oxidized flavoprotein + H(2)O.  
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U38218; AAA79722.1; -;  
 DR HSP; P00179; 1D76  
 DR InterPro: IPR001128; Cyt\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME P450; 1.  
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
 KW Microsome; Endoplasmic reticulum.  
 FT BINDING 443 443 HEME (BY SIMILARITY).  
 FT SEQUENCE 497 AA; 56010 MW; 3594AA88F04E58B1 CRC64;  
 SQ  
 Query Match 3.1%; Score 123.5; DB 1; Length 497;  
 Best Local Similarity 22.9%; Pred. No. 0.16;  
 Matches 104; Conservative 47; Mismatches 140; Indels 163; Gaps 22;  
 QY 343 TEAVLL-----GGWLFQHQPFQWGAQIRAVILGAVPRGRLVLDLFAESQPVTYRT 394  
 Db 12 TVAIFLLVLDLHRRQRWAARYPP---GPLPLPL-----GNLLHVD----- 50  
 QY 395 ASFGQGFPTWCMLEHFGN-----HGLFGALEA-VNGGPEAARL--FPNST 437  
 Db 51 --FKNTPYCFDQLRRFGNVSFLQAWTPVVLNGLAARVREALVTCGEDTADRPVPINQ 108  
 QY 438 MVGTGMAPEGI-----SQNEVYVSLMAELGWRKDPVDPDLAAWVTSFAARRYGVS 487  
 Db 109 VLGFGRSOGVFLARYGPAAWRQRFSVSTLNGLGK---KSLQWVTEEAACLCAPT 165  
 QY 488 PDAGAAWR---LLRSVYVNCSEACRGHNSPLVRRPSLQMTSIIWYNSRSDVFEAWRL 544  
 Db 166 DQGRFRPNSLLDKAVSN-----VIASLTGR----- 193  
 QY 545 TSAPSLATSPAFRYD-----LLDLTROAVQELVSLVYEEARSAYLSKELAS---LLR-- 593  
 Db 194 -----REYDDPRLRLFDLTALKE-----ESGLFRELVAIPAIPLLLRIP 234  
 QY 594 --AGGVLAIE--LLPALDEVL-----ASDSRFLGLSWLEQARAAAVSEAEADFYEQ 640  
 Db 235 GLAGKVLRSQKAFATQLODELLTHEHRTWTPAQPPDLTEAFLEAEKAK-GNPSSFNEE 293  
 QY 641 NSRYOLTWGPBEGNILDVANKLAGLVANYTPRRLFLEALVDVSAQIPFQHQFDKN 700  
 Db 294 NLR-----WVADLFSAGMVTSTTLAWGLLLMLILHPDVQRRV---QOEIDV 338

QY 701 VFQLEQAFVLSKQRP-----SQPRGDTVDL 726  
 Db 339 IGQVRRPEMGDQARMPVYTTAVIHVQRFQGVPL 372  
 RESULT 3  
 ACVS\_NOCCLA  
 ID ACVS\_NOCCLA STANDARD; PRT; 3649 AA.  
 AC P27743;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Delta-(L-alpha-aminoadipyl)-L-cysteine-D-valine synthetase  
 DE (EC 6.-.-.-) (ACV synthetase) (ACVS).  
 GN PCBAB.  
 OS Noccardia lactamdurans.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;  
 CC Anycolatopsis.  
 OX NCBI\_TaxID=1913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VAR LC 411;  
 RX MEDLINE=92065808; PubMed=1956290;  
 RA Coque J.J.R., Martin J.F., Calzada J.G., Liras P.;  
 RT multidomain peptidase synthetase, and pcbaB, encoding a large  
 RT clustered together in an organization different from the same genes  
 RT in Acromonium chrysogenum and Penicillium chrysogenum.";  
 RL Mol. Microbiol. 5:1125-1133(1991).  
 CC -!- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE  
 CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS  
 CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER  
 CC INTERMEDIATES.  
 CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.  
 CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND  
 CC CEPHALOSPORIN.  
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME  
 CC FAMILY.  
 CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X57310; CAA40561.1; -;  
 DR PIR; S18268; S18268.  
 DR HSP; P14687; 1AMU.  
 DR InterPro: IPR000873; AMP-bind.  
 DR InterPro: IPR001242; DUF4.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR InterPro: IPR003880; Phosphopant\_attach.  
 DR InterPro: IPR001031; Thioesterase.  
 DR Pfam; PF00501; AMP-binding; 3.  
 DR Pfam; PF00668; Condensation; 3.  
 DR Pfam; PF00550; pp-binding; 3.  
 DR Pfam; PF00975; Thioesterase; 1.  
 DR PRINTS; PR00154; AMPBINDING.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.  
 DR PROSITE; PS00455; AMP-BINDING; 1.  
 DR PROSITE; PS00075; ACP\_DOMAIN; 3.  
 DR Ligase; Antibiotic biosynthesis; Multifunctional enzyme;  
 DR Repeat; Phosphopantetheine.  
 KW REPEAT 401 861 DOMAIN 1 (ADIPATE-ACTIVATING).  
 FT REPEAT 1014 1937 DOMAIN 2 (CYSTEINE-ACTIVATING).  
 FT REPEAT 2079 2985 DOMAIN 3 (VALINE-ACTIVATING).  
 FT DOMAIN 788 857 ACYL CARRIER (ACP) 1.  
 FT DOMAIN 1864 1933 ACYL CARRIER (ACP) 2.  
 FT DOMAIN 2910 2981 ACYL CARRIER (ACP) 3.

|  |                                |   |      |                                      |  |
|--|--------------------------------|---|------|--------------------------------------|--|
| FT   | BINDING                        | 820   | 820  | PHOSPHOPANTHETHEINE (BY SIMILARITY). |  |
| FT   | BINDING                        | 1896  | 1896 | PHOSPHOPANTHETHEINE (BY SIMILARITY). |  |
| FT   | BINDING                        | 2944  | 2944 | PHOSPHOPANTHETHEINE (BY SIMILARITY). |  |
| FT   | ACT_SITE                       | 3502  | 3502 | THIOESTERASE (BY SIMILARITY).        |  |
| SQ   | SEQUENCE                       | 3649 AA; 404079 MW; 6FD095704F858E6B CRC64;                   |      |                                      |  |
| Query Match 3.0%; Score 119; DB 1; Length 3649;                  |                                |   |      |                                      |  |
| Best Local Similarity 21.1%; Pred. No. 4.9; Indels 306; Gaps 37; |                                |   |      |                                      |  |
| Matches 174; Conservative 63; Mismatches 283;                    |                                |   |      |                                      |  |
| QY   | 16                             | AGAGGAAGDEAEAAARALV-----ARLLGPGAADEFSVERALAAKPGLDYSLGG        | 69   |                                      |  |
| DB   | 1988                           | AALPGALGLTVRRHPALRTLLKTDGVRQYPIPADVRLEVP-----                 | 2031 |                                      |  |
| QY   | 70                             | GGAARVRVSGTGVAAGLHRYLDFCCCHVANGSSQRLRPPLPAVPGELTEATPNRY       | 129  |                                      |  |
| DB   | 2032                           | -----STTVDSRAELDEVLT-----RAGYVFLRHEELPI-----RAEAFDHGD         | 2070 |                                      |  |
| QY   | 130                            | RYQNVCTQSYFVWMDWAREREI-----DWMALNGINLAWSGQEAIWQRYV            | 179  |                                      |  |
| DB   | 2071                           | EIYLSVVVHSCDQNSWDFIRELAALLDGVPEADLGRGT-----YGEFAVWQRY         | 2124 |                                      |  |
| QY   | 180                            | LALGLTQAEINEFTGPAFLAWGMGNLHT-----WDGPLPPSWHIK-----QYLYLQHRVLD | 231  |                                      |  |
| DB   | 2125                           | LT-GKRLAALTEFWT-----GALGGFETIALPLDHPRPFRDYGRELEFELDERTTE      | 2176 |                                      |  |
| QY   | 232                            | QMSRFGMT-----   | 239  |                                      |  |
| DB   | 2177                           | ALRELARTARVSYSVLLGAWLMLNNYTGQDLVLVGTSPSANGRPEDFRANGFFANLLA    | 2236 |                                      |  |
| QY   | 240                            | -----PVLPAFAGHVPNAVTRVFPQVNVTRKMSWGFHCNCSYCSFLIAPEDPI-FPI     | 290  |                                      |  |
| DB   | 2237                           | LRVRVDPAAITLPAVRSVGEAV-VAQVH-----GELPFQLVKELKVEKDPDSRHPI      | 2287 |                                      |  |
| QY   | 291                            | IGSLFRELKEFGTDHIYGADTFNEMQPPSPSYLAATTAIVYEAAMTAVDTEAVWLQ      | 350  |                                      |  |
| DB   | 2288                           | LQLNFTLQNV-----SDH---TSALTGYQDPS-----GGWTTTKFD-LSATMTETATGLA  | 2333 |                                      |  |
| QY   | 351                            | GWL-----FOHQPFMGPAQIRAVLGAVPRGRLLVLDLFAESQ-PVY                | 391  |                                      |  |
| DB   | 2334                           | GNLTYAASLEDDTSASGFIATFRHV-----LAEFASAAAGTPIAQLTALDEPGQAALPDA  | 2388 |                                      |  |
| QY   | 392                            | TRTASFQGFQFIMCLMHFGNHLGALFAVNGGPEARLFNSTWVGTF-----MAPEGI      | 448  |                                      |  |
| DB   | 2389                           | TRRARRPGP-----GRCTRLF-----EEVAATWPDRAVAVHGDVRLTYREL           | 2430 |                                      |  |
| QY   | 449                            | SQ-NEVYISLMAGLWRKDP-----VPDLAAWVTSFAARRYGVSHPDAGAA            | 493  |                                      |  |
| DB   | 2431                           | NERANRLAHLRSVAEPRADELIAVLVDKSELTLVAIILAVWKAGAAIYMPIDPSTPDDRIA | 2490 |                                      |  |
| QY   | 494                            | W-----RLLRSVYNCSGACRGNHRSPLVRRPSLQMNTSIWYNSRDVFEAWRLLLTS      | 546  |                                      |  |
| DB   | 2491                           | FMLSSTGAKVLV-----AGEAHSVRG-----LTSGDVLDLQLDLTG                | 2529 |                                      |  |
| QY   | 547                            | APSLATSPAFRYDLDLTRQAVQELSVLYEARSAYLSKELASLLRAGGVLAYELLPAL     | 606  |                                      |  |
| DB   | 2530                           | EP-AENP-----VTETSTELAVAIY---TSGTTGPKPAVLVSHGVS-----           | 2567 |                                      |  |
| QY   | 607                            | DEVLASDRFLLGSLWLEOARAAAEAEADFEYQNSRVLTLWGPENIL-----           | 656  |                                      |  |
| DB   | 2568                           | DSFRAQLSGRYFGSPDESAEAVLFL---ANVVFDSVEQLALSVLGKHKLVLPPPSAADD   | 2624 |                                      |  |
| QY   | 657                            | ----DYANKQLAGLVANYITPRWRLEFLFALVDLSVAQGIPIFQHQHFD             | 698  |                                      |  |
| DB   | 2625                           | PAFYELANRE-GL-----SYLSGTPPTQVERFD                             | 2650 |                                      |  |
| RESULT   | 4                              |   |      |                                      |  |
| ID   | CPD6_HUMAN                     |   |      |                                      |  |
| AC   | P10635; O16752;                | STANDARD;   | PRT; | 497 AA.                              |  |
| DT   | 01-JUL-1989 (Rel. 11, Created) |   |      |                                      |  |

|    |  |  |
|----|--|--|
| DT | 01-JUL-1989 (Rel. 11, Last sequence update)                            |  |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update)                          |  |
| DE | Cytochrome P450 2D6 (EC 1.14.14.1) (CYP11D6) (P450-DB1) (Debrisoquine  |  |
| DE | 4-hydroxylase).  |  |
| GN | CYP2D6.  |  |
| OS | Homo sapiens (Human).  |  |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;                |  |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.             |  |
| OX | NCBI_TaxID=9606;   |  |
| RN | [1]  |  |
| RP | SEQUENCE FROM N.A.   |  |
| RX | MEDLINE=86314103; PubMed=3410476;                                      |  |
| RA | Gonzalez F.J., Vilbois F., Hardwick J.P., McBride O.W.,                |  |
| RA | Gonzalez D.W., Gelboin H.V., Meyer U.A.;                               |  |
| RT | "Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino   |  |
| RT | acid sequence and assignment of the CYP2D locus to chromosome 22.";    |  |
| RL | Genomics 2:174-179(1988).  |  |
| RN | [2]  |  |
| RP | SEQUENCE FROM N.A.   |  |
| RC | TISSUE=Liver;  |  |
| RX | MEDLINE=88122614; PubMed=3123997;                                      |  |
| RA | Gonzalez F.J., Skoda R.C., Kimura S., Umeno M., Zanger U.M.,           |  |
| RA | Nebert D.W., Gelboin H.V., Hardwick J.P., Meyer U.A.;                  |  |
| RT | "Characterization of the common genetic defect in humans deficient in  |  |
| RT | debrisoquine metabolism.";   |  |
| RL | Nature 331:442-446(1988).  |  |
| RN | [3]  |  |
| RP | SEQUENCE FROM N.A.   |  |
| RX | MEDLINE=90072069; PubMed=2574001;                                      |  |
| RA | Kimura S., Umeno M., Skoda R.C., Meyer U.A., Gonzalez F.J.;            |  |
| RT | "The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and      |  |
| RT | identification of the polymorphic CYP2D6 gene, a related gene, and a   |  |
| RT | pseudogene.";  |  |
| RL | Am. J. Hum. Genet. 45:889-904(1989).                                   |  |
| RN | [4]  |  |
| RP | VARIANT LYS-281 DEL (CYP2D6*9).  |  |
| RX | MEDLINE=93244880; PubMed=1844820;                                      |  |
| RA | Tyndale R., Aoyama T., Broly F., Matsunaga T., Inaba T., Kalow W.,     |  |
| RA | Gelboin H.V., Meyer U.A., Gonzalez F.J.;                               |  |
| RT | "Identification of a new variant CYP2D6 allele lacking the codon       |  |
| RT | encoding Lys-281: possible association with the poor metabolizer       |  |
| RT | phenotype.";   |  |
| RL | Pharmacogenetics 1:26-32(1991).  |  |
| RN | [5]  |  |
| RP | VARIANT SER-34 AND THR-486 (CYP2D6*10).                                |  |
| RX | MEDLINE=94115362; PubMed=8287064;                                      |  |
| RA | Yokota H., Tamura S., Furuya H., Kimura S., Watanabe M., Kanazawa I.,  |  |
| RA | Kondo I., Gonzalez F.J.;   |  |
| RT | "Evidence for a new variant CYP2D6 allele CYP2D6J in a Japanese        |  |
| RT | population associated with lower in vivo rates of sparteine            |  |
| RT | metabolism.";  |  |
| RL | Pharmacogenetics 3:256-263(1993).                                      |  |
| RN | [6]  |  |
| RP | VARIANT PRO-324 (CYP2D6*7).  |  |
| RX | MEDLINE=95147995; PubMed=7845481;                                      |  |
| RA | Evert B., Griese E.U., Eichelbaum M.;                                  |  |
| RT | "A missense mutation in exon 6 of the CYP2D6 gene leading to a         |  |
| RT | histidine 324 to proline exchange is associated with the poor          |  |
| RT | metabolizer phenotype of sparteine.";                                  |  |
| RL | Naunyn-Schmiedeberg Arch. Pharmacol. 350:434-439(1994).                |  |
| RN | [7]  |  |
| RP | VARIANT GLU-212 (CYP2D6*6B/6C).  |  |
| RX | MEDLINE=95172594; PubMed=7868129;                                      |  |
| RA | Daly A.K., Leathart J.B., London S.J., Idle J.R.;                      |  |
| RT | "An inactive cytochrome P450 CYP2D6 allele containing a deletion and a |  |
| RT | base substitution.";   |  |
| RL | Hum. Genet. 95:337-341(1995).  |  |
| RN | [8]  |  |
| RP | VARIANT ILE-107 (CYP2D6*17).   |  |
| RX | MEDLINE=97126511; PubMed=8971426;                                      |  |
| RA | Masimirembwa C., Persson I., Bertilsson L., Hasler J.,                 |  |
| RA | Ingelman-Sundberg M.;  |  |
| RT | "A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a     |  |





```

QY 707 -----AFVLSKQRYPSQPRGDTVDL 726
Db 351 AHMPYTTAVIHEVQRP-----GDIVPL 372

RESULT 5
YK95_RHIME STANDARD; PRT: 1029 AA.
AC Q52999;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein R02095.
GN R02095 OR SMC01491.
OS Rhizobium melliloti (Sinorhizobium melliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Anpe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puhler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallibert F.,
RA "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium melliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
RN [2]
RP SEQUENCE OF 1-483 FROM N.A.
RC STRAIN=1021;
RA Bent A.F., Signer E.R.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 GUANYLATE CYCLASE DOMAIN.
CC -1- SIMILARITY: TO M.LEPRAE MLCB2407.09.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
CC FRAMESHIFTS IN POSITIONS 311 AND 321.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL591789; CAC46674.1; -
DR EMBL; M30934; AAA88525.1; ALT_FRAME.
DR InterPro; IPR001054; Guanylt_cyclase.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW Hypothetical protein; ATP-binding; Complete proteome.
FT DOMAIN 39 168 GUANYLATE CYCLASE.
FT NP_BIND 261 323 ATP (POTENTIAL).
FT CONFLICT 323 323 A -> P (IN REF. 2).
SQ SEQUENCE 1029 AA; 112709 MW; 4EA478188B601958 CRC64;

Query Match 2.8%; Score 110.5; DB 1; Length 1029;
Best Local Similarity 19.0%; Pred. No. 3.7;
Matches 169; Conservative 98; Mismatches 279; Indels 343; Gaps 42;

QY 17 GAGGAGDGAREAAAVARLVARLGGP-----AADFVSVERAALAAKPGLD 63
Db 26 GGSVVRGGER---IVTALCYDLVGSITDLMHVMDIEDYQELMSAFQLASKQAISHSGVM 82
QY 64 TYSLGGGGAA-----RVVRVSGTGVAAAAGLHRYLRDFC----- 97
Db 83 QHEAGDGGVALPPIELKADAASLAIRAGL--IVEACKRVGREAGQDLDQVRVGIATSV 140

```

comparison of rhsA with other members of the rhs multigene family. ";  
 [2]  
 J. Bacteriol. 172:446-456(1990).  
 RC SEQUENCE FROM N.A.  
 RX STRAIN=K12 / MG1655;  
 RA MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes.";  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 RN [3]  
 RP REVIEW  
 RX MEDLINE=95020608; PubMed=7934896;  
 RA Hill C.W., Sandt C.H., Vlazny D.A.;  
 RT "Rhs elements of Escherichia coli: a family of genetic composites  
 RT each encoding a large mosaic protein.";  
 RL Mol. Microbiol. 12:865-871(1994).  
 CC -!- FUNCTION: RHS ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY  
 CC PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.  
 CC -!- DOMAIN: EACH RHS APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 kDa  
 CC AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.  
 CC -!- SIMILARITY: BELONGS TO THE RHS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: L19044; AAC95065.1; -;  
 DR EMBL: U00039; AAB18570.1; -;  
 DR EMBL: AE000437; AAC76617.1; -;  
 DR EcoGene; EG10846; rhsA.  
 DR InterPro; IPR001826; RHS.  
 DR PRINTS; PR00394; RHSPROTEIN.  
 KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1377  
 FT TRANSMEM 28 55  
 FT DOMAIN 330 1186  
 FT REPEAT 330 352  
 FT REPEAT 353 374  
 FT REPEAT 375 417  
 FT REPEAT 418 438  
 FT REPEAT 439 460  
 FT REPEAT 461 481  
 FT REPEAT 482 502  
 FT REPEAT 503 525  
 FT REPEAT 526 546  
 FT REPEAT 547 567  
 FT REPEAT 568 588  
 FT REPEAT 589 609  
 FT REPEAT 610 629  
 FT REPEAT 630 650  
 FT REPEAT 651 671  
 FT REPEAT 672 691  
 FT REPEAT 692 711  
 FT REPEAT 712 734  
 FT REPEAT 735 758  
 FT REPEAT 759 788  
 FT REPEAT 808 829  
 FT REPEAT 829 850  
 FT REPEAT 851 871  
 FT REPEAT 872 894  
 FT REPEAT 895 930  
 FT REPEAT 931 959  
 FT REPEAT 960 984  
 FT REPEAT 985 1019  
 FT REPEAT 1162 1186  
 SQ SEQUENCE 1377 AA; 156320 MW; 21ACA989E74200FE CRC64;

Query Match 2.8%; Score 109.5; DB 1; Length 1377;  
 Best Local Similarity 20.8%; Pred. No. 6.5;  
 Matches 128; Conservative 58; Mismatches 209; Indels 219; Gaps 35;  
 QY 57 AAKPGLDYSLGGG--GAARVRVSGSTGVAAA---AGL--HRYLRDFCGCHVAVSGSQL 108  
 DB 6 AARQG-DMTOYGGSIYQGSAGVRICAGTGVACVPGGVGTSGHPVNPILLGAKVLPGETDI 64  
 QY 109 RLPRPLPAV-----PG-----ELTEAT-----PNRYRYQ 133  
 DB 65 ALPGPLPFLTSRTSYSTRTKTPAPVGLPGWKMFPADIRLQLRDNTLILSDNGGSLIFE 124  
 QY 134 NVC--TQSYFVWWDWAREREIDWALNGINLAWGQEARIMORV-----YLA 181  
 DB 135 HLFPGEDGYS-----RSESLWLVGGVAKLDEGHRLAALWQALPEELRSLSPHYLA 175  
 QY 182 LGLTQAEINEFTGPFLAWGRMNLHTW-----DGPLPPSHHIQVLYLQHRVL 230  
 DB 176 TNSPQ-----GP-----WLLG-----WCERVPEADVLPALEP-----YRVL 209  
 QY 231 DOM-RSFGMTVPVLPFAGHVEAVTRVPQVNTKMSWGHFNCSYSCSFLAPEDPIFP 289  
 DB 210 TGLVDRFGRTQTFHR-----EAAGEFSGEITGVTGAWRHFRLVLTQAQRAEE----- 258  
 QY 290 IIGSLFLRELKKEFGTDHIYGADTFENMQPSSPSYLAATTAAYEAMTAVDTAEVWLL 349  
 DB 259 -----ARQQAISGGE-----PSAFPDTLPGYTE--YGRDNGIRLSAVWL- 296  
 QY 350 QGWLFQHPQFQWGAQIRAV-----LGAVPRGRLLVLDLFAESQPVYTRTASFQGPFTWC 405  
 DB 297 -----THDPEY--PENLPAAPLVRYGWTGPRGELAV-----VYDRSGK----- 331  
 QY 406 MLHNEGGNHGLFGALEA--VNGGPEAARLFNSTWVTGMAPEGIS----- 449  
 DB 332 QVRSFTYDDKYRGRVVAHRHTGRPEIRYRYSDSGRVTEQLNPAGLSYTYQEKDRITTD 391  
 QY 450 ---QNEVYSLMAELGWRK-----DVPDLAAWVTSFAARRYGVSHPD 490  
 DB 392 SLDRREVLHT--QGEAGLKRVRVKEHADGVSQSFDAVGLRLA--QTDAGRTEYSPDV 448  
 QY 491 GAANRLLLRVYVNCSEAC-----RGHNRSLVRRP-SLOWNTSIWNRSDVFEARLLLT 545  
 DB 449 VTG---LITRITTPDGRASAFYNNHNLQTSATPGDGL------RREYDELGRLIQE 498  
 QY 546 SAPSLATSPAFRYD 559  
 DB 499 TAPD-GDITRYRD 511  
 RESULT 7  
 RHSB\_ECOLI  
 ID RHSB\_ECOLI STANDARD; PRT; 1411 AA.  
 AC P16917; P76701;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RhsB protein precursor.  
 GN RHSB OR B3482.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93259920; PubMed=8387990;  
 RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;  
 RT "Rhs elements of Escherichia coli K-12: complex composites of shared  
 RT and unique components that have different evolutionary histories.";  
 RL J. Bacteriol. 175:2799-2808(1993).  
 RN [2]  
 RP REVISION TO 405.  
 RA Hill C.W.;

Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

[3]  
SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RA MEDLINE-94316500; PubMed-8041620;  
RX Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
"Analysis of the Escherichia coli genome. V. DNA sequence of the  
region from 76.0 to 81.5 minutes."  
RT Nucleic Acids Res. 22:2576-2586(1994).  
[4]  
SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
"The complete genome sequence of Escherichia coli K-12."  
RT Science 277:1453-1474(1997).  
[5]  
SEQUENCE OF 1-100 FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-89123133; PubMed-2644231;  
RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;  
"rbs gene family of Escherichia coli K-12."  
RT J. Bacteriol. 171:636-642(1989).  
[6]  
SEQUENCE OF 1221-1411 FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-90094253; PubMed-2403547;  
RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,  
Vlazny D.A., Zhang J., Zhao S., Hill C.W.;  
"Structure of the rbsA locus from Escherichia coli K-12 and  
comparison of rbsA with other members of the rbs multigene family."  
RT J. Bacteriol. 172:446-456(1990).  
[7]  
REVIEW.  
RX MEDLINE-95020608; PubMed-7934896;  
RA Hill C.W., Sandt C.H., Vlazny D.A.;  
"Rbs elements of Escherichia coli: a family of genetic composites  
each encoding a large mosaic protein."  
RT Mol. Microbiol. 12:865-871(1994).  
RL FUNCTION: RBS ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY  
PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.  
CC -1- DOMAIN: EACH RBS APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 kDa  
AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.  
CC -1- SIMILARITY: BELONGS TO THE RBS FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L02370; AAC61883.1; -;  
DR EMBL; U00039; AAB18457.1; -;  
DR EMBL; AE000424; AAC76507.1; -;  
DR EcoGene; EG10847; rbsB.  
DR InterPro; IPR001826; RBS.  
DR PRINTS; PR00394; RBSPROTEIN.  
KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
FT SIGNAL 1 26  
FT CHAIN 27 1411  
FT RBSB PROTEIN.  
FT POTENTIAL.  
FT DOMAIN 330 1186  
FT REPEAT 330 352  
FT REPEAT 353 374  
FT REPEAT 375 417  
FT REPEAT 418 438  
FT REPEAT 439 460  
FT REPEAT 461 481

FT REPEAT 482 502 7.  
FT REPEAT 503 525 8.  
FT REPEAT 526 546 9.  
FT REPEAT 547 567 10.  
FT REPEAT 568 588 11.  
FT REPEAT 589 609 12.  
FT REPEAT 610 629 13.  
FT REPEAT 630 650 14.  
FT REPEAT 651 671 15.  
FT REPEAT 672 691 16.  
FT REPEAT 692 711 17.  
FT REPEAT 712 734 18.  
FT REPEAT 735 758 19.  
FT REPEAT 808 828 20.  
FT REPEAT 829 850 21.  
FT REPEAT 851 871 22.  
FT REPEAT 872 894 23.  
FT REPEAT 895 930 24.  
FT REPEAT 931 959 25.  
FT REPEAT 960 984 26.  
FT REPEAT 985 1019 27.  
FT REPEAT 1162 1186 28.  
FT CONFLICT 1130 1130 Q -> K (IN REF. 4).  
SQ SEQUENCE 1411 AA; 159394 MW; 567AC4EE713D9E07 CRC64;  
  
Query Match 2.8%; Score 109.5; DB 1; Length 1411;  
Best Local Similarity 20.8%; Pred. No. 6.7;  
Matches 128; Conservative 58; Mismatches 209; Indels 219; Gaps 35;  
  
QY 57 AAKPGLDYSLLGGG---GAARVRVGRSTGVAAA---AGL--HRYLRDFCGCHVAMSGSQL 108  
DB 6 AARQG-DMTQYGGSIQVSAGVRICAPGVACSVCPGGVTSGHPNPLIGAKVLPGETDI 64  
QY 109 RLPRPLPAV-----PG-----PG-----ELTEAT-----PNRYRYQ 133  
DB 65 ALPGPLPFLISRTYSYRTKTPAPVGLPGWKMPADIRLQLRONTLILSDNGGRSLYFE 124  
QY 134 NVC--TQSYFVWDWARKERIDWALNGINLALWSGOEAIQRV-----YLA 181  
DB 125 LGLTDAEINEFFTGAFLAWGMGNLHTW-----RSESLWLRGVAKLDEGHRLAALWQALPEELKLSPHRYLA 175  
QY 182 LGLTDAEINEFFTGAFLAWGMGNLHTW-----RSESLWLRGVAKLDEGHRLAALWQALPEELKLSPHRYLA 230  
DB 176 TNSPQ-----GP-----WLLG---WCEVRPEADEVLPAFLPP-----YRVL 209  
QY 231 DQM-RSFGMTPLVLPAPAGHVPPEAVTRVFPQVNVTKMGSHGHCNCSYCSFLLAPEDPIFP 289  
DB 210 TGLVDRFGRQTQFHR-----EAAGEFSGEITGVTDGAWRHFLRLVLTQQAQRAEE-----258  
QY 290 IIGSLFLRELKEFGDTHIYGADTFNEMOPSPSEPSYLAATAATVAYEAMTAVDTAEVWLL 349  
DB 259 -----ARQAISGGTE-----PSAFDPTLPGYTE--YGRDNGIRLSAVWL- 296  
QY 350 QGWLFOQFOQFMGPAQIRAV---LCAVPRGRLLVLDLFAESQPVYVTRTASFQGGPFIMC 405  
DB 297 -----THDPEY--PENLPAALVRYGWTGPRGELAV-----VYDRSGK-----331  
QY 406 MLHNFNGNIGLFGALEA--VNGGPEAARLPFNSTWGTGMAPEGIS-----449  
DB 332 QVRSFTYDDKYRGMVVAHRHTGRPEIRYRSDGRVTEQLNPAGLSYTYQYEKDRITTYD 391  
QY 450 ---QNEVWYSLMAELGWRK-----DPVPDLAAVWTSFAARRYGVSHPDA 490  
DB 392 SLDRREVLHT-QGEAGLKRVRKKEHADGSGVTSQSFDAVGLRLA--QTDAAGRTTEYSPOV 448  
QY 491 GAARLRLLSVNCNCSGEAC-----RGNHRSPLVRRP-SLOMNTSIWYNRSDVFEAWRLILT 545  
DB 449 VTG---LITRITTPDGRSAFYNNHNNQLTSATGPDGLEI-----RREYDELGLIOE 498  
QY 546 SAPSLATSPAFRYD 559  
DB 499 TAPD-GDITRYRYD 511

|    |          |         |           |                       |
|----|----------|---------|-----------|-----------------------|
| FT | CONFLICT | 233     | 233       | D -> N (IN REF. 1).   |
| FT | CONFLICT | 275     | 275       | A -> P (IN REF. 1).   |
| FT | CONFLICT | 545     | 545       | G -> R (IN REF. 1).   |
| FT | CONFLICT | 549     | 549       | S -> H (IN REF. 1).   |
| FT | CONFLICT | 553     | 553       | A -> P (IN REF. 1).   |
| FT | CONFLICT | 561     | 561       | Q -> H (IN REF. 1).   |
| FT | CONFLICT | 634     | 634       | R -> K (IN REF. 1).   |
| SQ | SEQUENCE | 726 AA; | 73656 MW; | 7C4BA44395FFAB CRC64; |

  

|                       |  |                  |                 |             |             |
|-----------------------|--|------------------|-----------------|-------------|-------------|
| Query Match           |  | 2.7%;            | Score 108;      | DB 1;       | Length 726; |
| Best Local Similarity |  | 17.7%;           | Pred. No. 3.5;  |             |             |
| Matches 134;          |  | Conservative 73; | Mismatches 196; | Indels 352; | Gaps        |

  

|    |     |            |               |          |            |            |            |                      |                         |                   |              |           |             |      |   |         |
|----|-----|------------|---------------|----------|------------|------------|------------|----------------------|-------------------------|-------------------|--------------|-----------|-------------|------|---|---------|
| Qy | 86  | AAGLHRYLRF | CGCHVAMSG     | SQRLRPL  | PAVPGLTEAT | PNR-----   | YRYQNVCTQ  | 138                  |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      | :: ::         | :: ::    | :: ::      | :: ::      | :: ::      |                      |                         |                   |              |           |             |      |   |         |
| Db | 26  | SAGAGTASRD | -----         | WPNQLRV  | -----      | DLLNQHSNRS | PLGDEFDYK  | FSKL 69              |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               | :: ::    |            | :: ::      |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 139 | SYSFV      | -----         | WWDWARWE | REIDWMA    | LNGINLALAS | -----      | 169                  |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               | :: ::    |            | :: ::      |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 70  | DYSALKGDL  | KALLTDSQPW    | -----    | PADWGSYVGL | FIRMAWHG   | AGTYRSIDGR | GAGR 122             |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 170 | -----      | -----         | -----    | -----      | GOEAIWQ    | RYVYALGLT  | QAEINEFFGPAF 198     |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      | :: ::      |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 123 | GOORFAPL   | NSWPNVSLDKARR | LLWPTIK  | QYGGKIS    | WADLFILAG  | NVALENG    | SFRT---- 178         |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      | :: ::      |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 199 | LAWGRMGN   | LHTWGP        | LPSPWHIK | OLYQHRV    | LQDMRSFG   | MTPVLP     | FAFAGHVPEAVTRVP 258  |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               | :: ::    | :: ::      | :: ::      |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 179 | FGFG       | -AGREDW       | EPDLV    | NMGDEKAW   | LTHR-----  | -----      | H-PEALAKA-- 215      |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 259 | QVNVTKMS   | GNHNCYS       | SCSFLI   | APEDPFP    | ITIGSLF    | LBELIKE    | FTDTHIYGADTFNEMQ 318 |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               | :: ::    | :: ::      | :: ::      |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 216 | PLGATENG   | -----         | LIYVNP   | PEGP-----  | -----      | -----      | DH----- 234          |                         |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 319 | PPSPEPY    | LAAATTA       | VYEAMT   | ADVTDA     | EVWLLQ     | GLFHQHP    | QFWGPAQIRAVL         | GAVPRGL 378             |                   |              |           |             |      |   |         |
|    |     | :: ::      |               | :: ::    |            | :: ::      |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 235 | --SGEPL    | SAAAI         | IRATFG   | NMGMD      | ETVALI     | AG--       | CHTLGK               | THGAAAASHVGADP--- 285   |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 379 | LVLDL      | FAESOP        | VYTRT    | ASFQCG     | PFICM      | LHFNCG     | NHGLF                | GALEAVNGGPEAA-- 430     |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      | :: ::      |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 286 | -----      | EAAPI         | -----    | EAOLGN     | WASSYGS    | -----      | VGA                  | -DAITSGLEVVMTQTPTQW 325 |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 431 | -----      | RLPSTN        | VMGTG    | MAPEGIS    | ONEVYVS    | -----      | -----                | ----- 466               |                   |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 326 | SNYFFEN    | LFXEW         | -----    | QTSR       | PAGAIQ     | EA         | VDAPDI               | DPDFDSKRRKPKMLVTD       | TLTRFD 383        |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 467 | P-----     | VPDLAA        | AWTYS    | FAARRY     | GVSH       | PDAGAA     | AWRLLL               | RSVYVNC                 | SGEACRGNHRSPL 517 |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 384 | PEPEKIS    | RRRLND        | PQAFNE   | AFAW       | EKFL       | THRDM      | GPKARYI              | -----                   | ----- 423         |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 518 | VRPS       | LQMTS         | IWYNS    | DSDFE      | AWRLL      | LT         | SAPS                 | LATSPAF                 | RYDLDDI-----      | TRQAVOEL 571 |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 424 | --GPEV     | PKEDLI        | WOD----- | -----      | PLPQ       | PLYOPT     | -QEDII               | NKAAIA                  | ASGLSISEM 467     |              |           |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 572 | VSLYYEE    | ARSAYLS       | KE       | LASLL      | RAG-----   | -----      | -----                | -----                   | -----             | CVLAYELLPAL  | DEVIL 610 |             |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 468 | VSV-----   | -----         | AWAS     | -----      | ASTF       | RGDK       | RG                   | GANGAR                  | LALAPQR           | OWDYNAA      | ARVLP     | LEEIQ 517   |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Qy | 611 | ADSR       | RELLGS        | -----    | -----      | WLEQ       | ARAAA      | -----                | -----                   | -----             | -----        | -----     | VSEAEAD 636 |      |   |         |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |
| Db | 518 | KTINKA     | S             | LADI     | IIVLAG     | VWGIE      | QAAAAA     | AGVS                 | ISVP                    | FAPGR             | VDARQ        | ODT       | DIEM        | FSLE | P | IAD 577 |
|    |     | :: ::      |               |          | :: ::      |            |            |                      |                         |                   |              |           |             |      |   |         |

|  |  |
|--|--|
| DT   | 16-OCT-2001 (Rel. 40, Last annotation update)  |
| DE   | Peroxidase/Catalase HPI (EC 1.1.1.6) (Catalase-peroxidase)   |
| DE   | (Hydroperoxidase I).   |
| GN   | KATG OR B3942.   |
| OS   | Escherichia coli.  |
| OS   | Bacteriophage proteobacteria; gamma subdivision; Enterobacteriaceae.   |
| OC   | Escherichia.   |
| OX   | NCBI_TaxID=562;  |
| RN   | [1]  |
| RP   | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  |
| RX   | MEDLINE=88314956; PubMed=3045098;  |
| RA   | Triggs-Raine B.L., Doble B.W., Mulvey M.R., Sorby P.A., Loewen P.C.;   |
| RT   | "Nucleotide sequence of katG, encoding catalase HPI of Escherichia   |
| RT   | coli.";  |
| RL   | J. Bacteriol. 170:4415-4419(1988).   |
| RN   | [2]  |
| RP   | SEQUENCE OF 1-339 FROM N.A.  |
| RC   | STRAIN=K12 / MG1655;   |
| RX   | MEDLINE=93347969; PubMed=8346018;  |
| RA   | Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;  |
| RT   | "Analysis of the Escherichia coli genome. III. DNA sequence of the   |
| RT   | region from 87,2 to 89,2 minutes.";  |
| RL   | Nucleic Acids Res. 21:3391-3398(1993).   |
| RN   | [3]  |
| RP   | SEQUENCE OF 309-726 FROM N.A.  |
| RC   | STRAIN=K12 / MG1655;   |
| RX   | MEDLINE=94089392; PubMed=8265357;  |
| RA   | Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,  |
| RA   | Daniels D.L.;  |
| RT   | "Analysis of the Escherichia coli genome. IV. DNA sequence of the  |
| RT   | region from 89,2 to 92,8 minutes.";  |
| RL   | Nucleic Acids Res. 21:5408-5417(1993).   |
| CC   | - FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND   |
| CC   | BROAD-SPECTRUM PEROXIDASE ACTIVITIES.  |
| CC   | - CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.   |
| CC   | - GOFACOR: TWO PROTOHEME IX GROUPS PER TETRAMER.   |
| CC   | - SUBUNIT: HOMOTETRAMER.   |
| CC   | - INDUCTION: BY HYDROGEN PEROXIDE.   |
| CC   | - PTM: The N-terminus is blocked.  |
| CC   | - SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL  |
| CC   | PEROXIDASE/CATALASE SUBFAMILY.   |
| CC   | - SIMILARITY: CONTAINS 1 HEME REGULATORY MOTIF (HRM).  |
| CC   | -----  |
| CC   | This SWISS-PROT entry is copyright. It is produced through a collaboration   |
| CC   | between the Swiss Institute of Bioinformatics and the EMBL outstation  |
| CC   | the European Bioinformatics Institute. There are no restrictions on its  |
| CC   | use by non-profit institutions as long as its content is in no way   |
| CC   | modified and this statement is not removed. Usage by and for commercial  |
| CC   | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |
| CC   | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |
| CC   | -----  |
| DR   | EMBL; M21516; AAA24040.1; -  |
| DR   | EMBL; L19201; AAB03074.1; -  |
| DR   | EMBL; U00006; AAC43048.1; -  |
| DR   | EMBL; AE000468; AAC76924.1; -  |
| DR   | PIR; JS0140; CSECHP.   |
| DR   | PIR; S40885; S40885.   |
| DR   | HSSP; P48534; IAPX.  |
| DR   | SWISS-2DPAGE; P13029; COLI.  |
| DR   | EcoGene; EG10511; katG.  |
| DR   | InterPro; IPR002016; peroxidase.   |
| DR   | Pfam; PF00141; peroxidase; 1.  |
| DR   | PRINTS; PR00458; PEROXIDASE.   |
| DR   | PROSITE; PS00435; PEROXIDASE_1; 1.   |
| DR   | PROSITE; PS00436; PEROXIDASE_2; 1.   |
| DR   | Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;   |
| KW   | Complete proteome.   |
| FT   | ACT_SITE 102 102 BY SIMILARITY.  |
| FT   | ACT_SITE 106 106 BY SIMILARITY.  |
| FT   | BINDING 267 267 PROXIMAL HEME LIGAND (BY SIMILARITY).  |
| FT   | DOMAIN 15 20 HRM 1 (POTENTIAL).  |
| FT   | CONFLICT 621 621 A -> G (IN REF. 1).   |
| FT   | SEQUENCE 726 AA; 80023 MW; 24D32EBED5DE9BD6 CRC64;   |
| QY   | 86 AAGLHRYLRDFCGCHVAMSGSOLRLPRLPAVPGELTETATPNR-----YR----- 130   |
| DB   | 26 SAGAGTTTRD-----WPNQLRV-----DLLNQHSNRNPLGEDFYRKEFSKL 69  |
| QY   | 131 -YY-----QNVCTQSYSEFVMDWARWEREIDWALNGINLALAWS----- 169  |
| DB   | 70 DYYGLUKDKLALLTESQP--WW-----PADWGSYAGLFIRMAWHGAGTYRSIDGRGA 120   |
| QY   | 170 -----GQEAIWQRYVYALGJUTQAEINFFFTGP 196  |
| DB   | 121 GRGQRFAPLNSWPDNVSILDKARRLLWPIKQYGGKISWADLFILAGNVALENSGFT-- 178   |
| QY   | 197 AFLANGRMGNLHTWDCPLPPSWHIKQLYLQHRVLDQMSFGMTVLPAPAGHVPEAVTRV 256   |
| DB   | 179 --FGFG-AGREDVWEPDLVDNMGDEKAWLTHR-----H-PEALAKA 215   |
| QY   | 257 FPQVNVTKMSGHFNCSYSCSFLAPDPIFPIIGSLFLRELIRKEFGTDHIYGADTFNE 316  |
| DB   | 216 --PLGATMG-----LIYVNEGP-----DH----- 234   |
| QY   | 317 MOPPSSEPSYLAATTAATVYEAATAVDTEAVWLQGLWLFQHQPFQWGPQAIRAVLGAIVRG 376  |
| DB   | 235 ----SGEPLSAAAIATATFGNMGMDDETVALIAGGHTLGKTHGAGPT---SNVGPDP-- 285  |
| QY   | 377 RLLVLDLFAESOPVYTR-----TASFGQOPFIWCMHLNFGGNH---GLF--- 417   |
| DB   | 286 -----EAAPIEEQGLGWASTVSGVGADATITSGLEVWVTQPTQWSNYFFENLFKY 336  |
| QY   | 418 -----GAL--EAVNGGPEAARLFPNSTWGTGMAPEGISQNEVYVSLMAELGNRKD 466  |
| DB   | 337 ENVQTRSPAGAIQFEAV-----DAPEIIPD-----PFDPSSKKRKTMLVTLTLPD 383  |
| QY   | 467 P-----VPDLAAWTSPFAARRGVSHPDAGAAWRLLLSRVNCSCEACRGNRSPL 517  |
| DB   | 384 PEFEKISRRFLNDPOAFNEAFARWPKFLTHROMGPKSRYI----- 423  |
| QY   | 518 VRPRLQMTSITWYNSRSDVFEAWRLLLTSAPSLATSPAFRYDLDDL-----TRQAVQEL 571  |
| DB   | 424 --GPEVPKEDLIWQD-----PLPOPIYNPT-EQDIIDLKFAIADSLGSVSEL 467   |
| QY   | 572 VSLYYEARSAYLSKELALLRAG-----GVLAYELLPALDELVL 610  |
| DB   | 468 VSV-----AWAS-----ASTFRGDKRGKGANGARLALMPQDMVDVNAARALPVLEKIQ 517   |
| QY   | 611 ASDSRELLGSLWLEQARAAAVSEA 633   |
| DB   | 518 KESGKASLADIIVLAGVGVGEKA 540  |
| RESULT 10  |  |
| P85B_RAT   | STANDARD; PRT; 722 AA.   |
| ID P85B_RAT  |  |
| AC Q63788;   |  |
| DT 15-DEC-1998 (Rel. 37, Created)  |  |
| DT 15-DEC-1998 (Rel. 37, Last sequence update)                           |  |
| DT 15-DEC-1998 (Rel. 37, Last annotation update)                         |  |
| DE Phosphatidylinositol 3-kinase regulatory beta subunit (PI3-kinase     |  |
| DE P85-beta subunit) (PtdIns-3-kinase P85-beta).                         |  |
| GN PIK3R2.   |  |
| OS Rattus norvegicus (Rat).  |  |
| OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;     |  |
| OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. |  |
| OX NCBI_TaxID=10116;   |  |
| RN [1]   |  |
| RP SEQUENCE FROM N.A.  |  |
| RC STRAIN=WISTAR; TISSUE=Brain;  |  |
| RX MEDLINE=96214979; PubMed=8621382;                                     |  |

Inukai K., Anai M., Vambreda E., Hosaka T., Katagiri H., Funaki M., Fukushima Y., Ogihara T., Yazaki Y., Kikuchi M., Oka Y., Asano T.:  
"A novel 55-kDa regulatory subunit for phosphatidylinositol 3-kinase structurally similar to p55PIK Is generated by alternative splicing of the p85alpha gene.";  
J. Biol. Chem. 271:5317-5320(1996).  
CC -|- FUNCTION: BINDS TO ACTIVATED (PHOSPHORYLATED) PROTEIN-TYROSINE KINASES, THROUGH ITS SH2 DOMAIN, AND ACTS AS AN ADAPTER, MEDIATING THE ASSOCIATION OF THE P110 CATALYTIC UNIT TO THE PLASMA MEMBRANE.  
CC -|- SUBUNIT: HETERODIMER OF A P110 (CATALYTIC) AND A P85 (REGULATORY) SUBUNITS.  
CC -|- SIMILARITY: CONTAINS 2 SH2 DOMAINS.  
CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -|- SIMILARITY: CONTAINS 1 RHO GAP DOMAIN.  
CC -|- SIMILARITY: HIGH, WITH OTHER P85-BETA SUBUNITS, AND ALSO TO P85-ALPHA SUBUNIT.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; D64046; BAA10926.1; --  
CC HSSP; P23727; 2PNB.  
CC InterPro: IPR001720; PI3K\_P85.  
CC InterPro: IPR000198; rhoGAP.  
CC InterPro: IPR000980; SH2.  
CC InterPro: IPR001452; SH3.  
CC Pfam; PF00620; rhoGAP; 1.  
CC Pfam; PF00017; SH2; 2.  
CC Pfam; PF00018; SH3; 1.  
CC PRINTS; PR00678; PI3KINASEP85.  
CC PRINTS; PR00401; SH2DOMAIN.  
CC SMART; SM00324; rhoGAP; 1.  
CC SMART; SM00252; SH2; 2.  
CC SMART; SM00326; SH3; 1.  
CC PROSITE; PS50001; SH2; 2.  
CC PROSITE; PS50002; SH3; 1.  
CC SH3 domain; SH2 domain; Repeat.  
FT DOMAIN 4 80 SH3.  
FT DOMAIN 125 257 RHO-GAP.  
FT DOMAIN 324 419 SH2 1.  
FT DOMAIN 616 710 SH2 2.  
SQ SEQUENCE 722 AA; 81328 MW; 1208368B9F6FC095 CRC64;

|                       |        |   |       |                                      |
|-----------------------|--------|---|-------|--------------------------------------|
| Query Match           | 2.6%;  | Score 103;  | DB 1; | Length 722;                          |
| Best Local Similarity | 21.6%; | pred. No. 8;  |       |                                      |
| Matches               | 87;    | Conservative  | 38;   | Mismatches 134; Indels 144; Gaps 19; |
| QY                    | 74     | RVRVSGTGVAAAGLHRYLRDFCGCHV-----AMSGSQLRPLRPAPVP-----        | 118   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 63     | RTRQRG-----DFEGTVFELGPVALARPGPRGRPLPARPLDGPSES              | 107   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 119    | ----GELTE-----ATPNRYRYQNVCTQSFSFVMW-----DNA-----RWER        | 152   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 108    | GHTLASLAEOFSPESAPPILVKLTIEAIOEALDSEFYSRPELPAPRTDMSLSLEQWDR  | 167   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 153    | EIDWMALINGINALAWSQEAIWRVVYLAGLTQAQNEFF-----TGPAFLAWGRMN     | 206   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 168    | TTLVDKVGCELLALP-----AAVVTPEASEAYRAMREVTPGVGLVLE----         | 210   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 207    | LHTWDGPLPPSWHKQLYLQHRVLDMQMSGMTPTLVLPFAFGHVPEAVTRYFPQNVTKMG | 266   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 211    | -----PPTLLHQALTURFLLOHLGRVARRAPSPATAVH---ALASAFGPL-----     | 253   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 267    | SWGHNCSYSCLSLPAD-----PIETPLGSFLRELKEPTGTHIYGADTFNE          | 316   |                                      |
| Db                    |        |   |       |                                      |
| QY                    | 254    | -----LLRAPPGCGGGSEPAPDFPV-----LLLERLVQEE-----HYDEQDTAPP     | 294   |                                      |
| Db                    |        |   |       |                                      |

```

QY 317 MOPPSSEPSYLAATAVYE-AMTAVDTFAVNLQGLWLFQHQPOFGWP-----AQTRAVLGA 372
Db 295 ALLEP--KPSKVKPAPATLANGSTPSLQDAEW-----YMGDISREEVNERLRD 340
QY 373 VPRGRLLYLDLFAESQPVYTRTASFQGPFTWCMLNHFGGNHG 415
Db 341 TPDGTFLYRDASSKIQGEYTLTLRKGNNKLIKVFHR-DGHYG 382

RESULT 11
YMCA_ECOLI
ID YMCA_ECOLI STANDARD; PRT; 698 AA.
AC P75882;
DC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein ymca precursor.
GN YMCA OR B0984.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]

```

```

SEQUENCE FROM N.A.
STRAIN:K12;
MEDLINE=97061202; PubMed=8905232;
RA Osima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
"A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
DNA Res. 3:137-155(1996).
-1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(CC (Probable).
-1- SIMILARITY: STRONG, TO E.COLI YJBR.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcements
or send an email to license@isb-sib.ch).
-----
CC EMBL; AS000200; AAC74069.1; -;
CC EMBL; D90735; BAA35749.1; -;
CC EMBL; D90736; BAA36124.1; -;
CC EcoGene; EG13729; ymcA.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Hypothetical protein; Membrane; Lipoprotein; Signal;
CC Complete proteome.
CC SIGNAL 1 18 POTENTIAL.
CC FT CHAIN 19 698 HYPOTHETICAL LIPOPROTEIN YMC A.
CC FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL) .
CC SEQUENCE 698 AA; 78687 MW; 37F71D2DC57FA4B7 CRC64;
SQ

```

Query Match 2.6%; Score 101; DB 1; Length 698  
Best Local Similarity 19.2%; Pred. No. 11;  
Matches 146; Conservative 75; Mismatches 239; Indels 3









QY 642 SRVQLTLCGP-----EGNILDYANKQLAGL-----VANYKTPRWR--- 676  
 DB 553 EMLRLTVGKTLQARVTAIEKYLKXDAQLNSWGCAPROVCHTTPVWVNDLSLTPDWNMT 612  
 QY 677 -----LFLALVDVAQGIPIFQHQDKNVFQLEQ 706  
 DB 613 WQEWKRVHYLEA---NISQLEQAQIQOEKNMYELQK 647

RESULT 14  
 NME4\_MOUSE  
 ID NME4\_MOUSE STANDARD; PRT: 1323 AA.  
 AC Q03391.  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl  
 D-aspartate receptor subtype 2b) (NR2D) (NMDAR2D).  
 GN GRIN2D.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93050214; PubMed=1395220;  
 RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,  
 RA Inoue Y., Mishina M.;  
 RT "Cloning and expression of the epsilon 4 subunit of the NMDA receptor  
 channel";  
 RL FEBS Lett. 313:34-38(1992).  
 RN [2]  
 RP REVISIONS.  
 RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,  
 RA Inoue Y., Mishina M.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS  
 CC POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT  
 CC SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.  
 CC -1- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D12822; BAA02254.1; -;  
 DR PIR: S27224; S27224.  
 DR HSSP: P19491; IGR2.  
 DR MGD: MGI:95823; Grin2d.  
 DR InterPro: IPR001320; Ion\_glut\_receptor.  
 DR InterPro: IPR001508; NMDA\_receptor.  
 DR InterPro: IPR001311; SBP\_glut\_receptor.  
 DR Pfam: PF00060; lig\_chan; 1.  
 DR PRINTS: PR00177; NMDARECEPTOR.  
 DR SMART: SM00079; PBPE; 1.  
 DR Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;  
 KW Ionic channel; Magnesium.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 1323 GLUTAMATE [NMDA] RECEPTOR SUBUNIT  
 FT EPSILON 4.  
 FT DOMAIN 28 581 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 582 601 1 (POTENTIAL).  
 FT DOMAIN 602 623 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 624 643 2 (POTENTIAL).  
 FT DOMAIN 644 653 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 654 672 3 (POTENTIAL).  
 FT DOMAIN 673 841 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 842 861 4 (POTENTIAL).  
 FT DOMAIN 863 1323 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 278 283 POLY-GLY.  
 FT DOMAIN 905 913 POLY-PRO.  
 FT DOMAIN 1030 1035 POLY-ALA.  
 FT DOMAIN 1197 1201 POLY-PRO.  
 FT SITE 639 FUNCTIONAL DETERMINANT OF NMDA  
 FT CARBOHYD 89 89 RECEPTORS (BY SIMILARITY).  
 FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1323 AA; 142907 MW; 8AE9878F90DD0921 CRC64;

Query Match 2.5%; Score 100; DB 1; Length 1323;  
 Best Local Similarity 22.7%; Pred. No. 30;  
 Matches 106; Conservative 46; Mismatches 156; Indels 158; Gaps 28;

QY 162 INLAWSQOEAIWQVYLALGLTOAEINEFTGPAFLAWGRMGLH-----TWDGGLP 215  
 DB 48 LNVALVFSGP-----AYAAEAARL-----GPAVAAVRSPLDVRPVVALVNGSDP 93  
 QY 216 PSMHIKQLYLQ-----HRVL--DQMRSEGMTVPVLPAPFAGHVPEAVTRVFPQVNT 263  
 DB 94 -----RSLVLQCLDLLSGLRVHGVFEDDSRAPAVAPILDFLSAQTSLPIVAVHG----- 143  
 QY 264 KMGSGHFNCSVCSFELLAPEDPIPIIGSLFLRELKIEFGTDHIYGADTFNMQ----- 318  
 DB 144 -----GAALVLTPEK-----GSTFLQ-----LGSSTEQQLGVIEV 175  
 QY 319 -PPSEPSYLAATTA-----VYEAMTAVDTEAVMLQGLWFOHQPF-WGPAQIRAVLGA 372  
 DB 176 LBEYDWTSEVATTPRAPHRAFLSIEVLTDG--SLVGW--EHRGALTLDPCAGEAVLGA 231  
 QY 373 VPRG-----RLVLVDLFAESQPVY--TRTASFOGQFFIWMCLH---NFGNGHGLFG-ALE 421  
 DB 232 QLRVSQAIRLLFC-AREAEVPFRAAEAGLTGPGYVMFVGPQLAGGSGGVGEPLL 290  
 QY 422 AVNGPEARLLPNSMTGVTGMAPEGISQNEVYSLMAELGWRKDPVDPDLAAWVTSFA-- 479  
 DB 291 LPGAPLPAGLF-----AVRSAGWRDRLARVAVAGVAVARG 327  
 QY 480 ----ARRYGVSHPDAGAWRLLLRSVYNGSGEACRHNRSPLVRPSLQ---MNTSIWYN 532  
 DB 328 AQALLRDYGLF-PEL-----GHDCRAQNRTH--RGESLHRYFNMNIT-WDN 368  
 QY 533 RSDVFEAWRLLTSPSLATSPAFRYDLDLIRQAVQELVSLYEE 578  
 DB 369 RDYSFNEDEGLVFN--PSLV-----VISLTRDRTWEVYVSWEQ 404

RESULT 15  
 NME4\_HUMAN  
 ID NME4\_HUMAN STANDARD; PRT: 1336 AA.  
 AC O15399;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl  
 D-aspartate receptor subtype 2b) (NR2D) (NMDAR2D) (EB11).  
 GN GRIN2D.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=98149394; PubMed=9489750;





|     |  |     |
|-----|--|-----|
| 125 | TPNRYRYQNVCTQSYSPFWWDWARWEIEDWMALNGINLALWSGGEATWQRYVYALGL  | 184 |
| QY  |  |     |
| 21  | TARRYRYQNVCTQSYSPFWWDWARWEIEDWMALNGINLALWSGGEATWQRYVYALGL  | 80  |
| Db  |  |     |
| 185 | TOAENIEFFTGFPAIAGRMGNLHTWDGPLPPSWHIKOLYLQHRVLDQMSFGMTPTVLP | 244 |
| QY  |  |     |
| 81  | TOAENIEFFTGFPAIAGRMGNLHTWDGPLPPSWHIKOLYLQHRVLDQMSFGMTPTVLP | 140 |
| Db  |  |     |
| 245 | FAGHVPEAVTRVFPQNVYTKMGSGHGFNCYSFSLAPEDPIFPIIGSLFLRELIIKEFG | 304 |
| QY  |  |     |

```
Db 141 FAGHVEAVTRFPQVNVTKMGSMGHNFCNSYSCSFLAPEDPIFFLIGSLFLRELKEEG 200
QY 305 TDHIYGADTFNEMOPSPSPSYLAATTAAYEAMTAVDTAEAVMLLQGWLFQHOPOFWGPA 364
Db 201 TDHIYGADTFNEMOPSPSPSYLAATTAAYEAMTAVDTAEAVMLLQGWLFQHOPOFWGPA 260
QY 365 QIRAVILGAVPRGRLVLDLFAESQPVYTRTASFQGPFTWCMHLNFGGHNGLFGALEAVN 424
Db 261 QIRAVILGAVPRGRLVLDLFAESQPVYTRTASFQGPFTWCMHLNFGGHNGLFGALEAVN 320
QY 425 GGPEARLFPNSTWGTGMAPEGISONEVYSLMAELGWRKDPVDPDLAAWVTSFAARRYG 484
Db 321 GGPEARLFPNSTWGTGMAPEGISONEVYSLMAELGWRKDPVDPDLAAWVTSFAARRYG 380
QY 485 VSHPDAGAAWRLLRSLVYNCSGEACRGNRSPVLRPRSLQMNNTSIWYNRSDVFEAWRLLL 544
Db 381 VSHPDAGAAWRLLRSLVYNCSGEACRGNRSPVLRPRSLQMNNTSIWYNRSDVFEAWRLLL 440
QY 545 TSAPSLATSPAFRYDLDLTLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLAELLP 604
Db 441 TSAPSLATSPAFRYDLDLTLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLAELLP 500
QY 605 ALDEVLASDRFLGLSLEQARAAVSEAEADFEQNSRYQLTLWGPENGLDYANKOLA 664
Db 501 ALDEVLASDRFLGLSLEQARAAVSEAEADFEQNSRYQLTLWGPENGLDYANKOLA 560
QY 665 GLVANYTTPRWRLFLDALVDSVAGQIPFQOHOQFDKNVFLQEQAFVLSKORYPSQPRGDTV 724
Db 561 GLVANYTTPRWRLFLDALVDSVAGQIPFQOHOQFDKNVFLQEQAFVLSKORYPSQPRGDTV 620
QY 725 DLAKKIFLKYPGWVAGSW 743
Db 621 DLAKKIFLKYPGWVAGSW 639

RESULT 2
ID O54752 PRELIMINARY; PRT; 739 AA.
AC O54752;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NAGLU (EC 3.2.1.50).
GN NAGLU.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAFLJ; TISSUE=LIVER;
RA Zhao K.W., Li H.H., Neufeld E.F.;
RT "Cloning and expression of Mouse gene encoding the lysosomal alpha-N-
RT acetylglucosaminidase.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85247; AAB8084.1; -.
DR MGD; MGI:1351641; Naglu.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
KW Hydrolyase; Glycosidase.
SQ SEQUENCE 739 AA; 82610 MW; B414C336A04EF0EF CRC64;

Query Match 83.9%; Score 3305; DB 11; Length 739;
Best Local Similarity 83.0%; Pred. No. 1.7e-236;
Matches 611; Conservative 53; Mismatches 70; Indels 2; Gaps 1;

QY 1 MEAVAAVAVGVLILAGAGAGAGDEARAAVRLVRLGLGPPAANDSVSVERALAKP 60
Db 1 MEAAGLAVILGLFLL--AGGSVGDEAREAKAVRELVRLLGLGPPAANFLVSVERALADES 58
QY 61 GLDITSLGGGGAARVVRGSGTGAAGLHRYLRDFCCGCHVAVSWSQLRPLRPVAPGPE 120
```

```
Db 59 GLDITSLGGGVPVLRGSGTGAAGLHRYLRDFCCGCHVAVSWSQLRPLRPVAPG 118
QY 121 LTEATPNRYRYQNVCTOSYSFVVMWDMARWEREDWMALNGINLALAWSGEATWQRYVL 180
Db 119 LTEATPNRYRYQNVCTOSYSFVVMWDMARWEREDWMALNGINLALAWSGEATWQRYVL 178
QY 181 ALGTLQAEINEFFTPAFPLAWGRMGNLHTWGPPLPPSHHIKOLYLQHRVLDQMRSGFWTP 240
Db 179 ALGTLQAEINEFFTPAFPLAWGRMGNLHTWGPPLPPSHHIKOLYLQHRVLDQMRSGFWTP 238
QY 241 VLPFAGHVPFAVTRFPQVNVTKMGSMGHNFCNSYSCSFLAPEDPIFFLIGSLFLREL 300
Db 239 VLPFAGHVPFAVTRFPQVNVTKMGSMGHNFCNSYSCSFLAPEDPIFFLIGSLFLREL 298
QY 301 KEFGTDHIYGADTFNEMOPSPSPSYLAATTAAYEAMTAVDTAEAVMLLQGWLFQHOPOF 360
Db 299 KEFGTDHIYGADTFNEMOPSPSPSYLAATTAAYEAMTAVDTAEAVMLLQGWLFQHOPOF 358
QY 361 WGPAGQIRAVILGAVPRGRLVLDLFAESQPVYTRTASFQGPFTWCMHLNFGGHNGLFGAL 420
Db 359 WGPAGQIRAVILGAVPRGRLVLDLFAESQPVYTRTASFQGPFTWCMHLNFGGHNGLFGAL 418
QY 421 EAVNGGPEARLFPNSTWGTGMAPEGISONEVYSLMAELGWRKDPVDPDLAAWVTSFAA 480
Db 419 EAVNGGPEARLFPNSTWGTGMAPEGISONEVYSLMAELGWRKDPVDPDLAAWVTSFAA 478
QY 481 RRYGVSHPDAGAAWRLLRSLVYNCSGEACRGNRSPVLRPRSLQMNNTSIWYNRSDVFEAW 540
Db 479 RRYGVSHPDAGAAWRLLRSLVYNCSGEACRGNRSPVLRPRSLQMNNTSIWYNRSDVFEAW 538
QY 541 RLLTSAPSLATSPAFRYDLDLTLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLA 600
Db 539 RLLTSAPSLATSPAFRYDLDLTLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLA 598
QY 601 ELLPALDEVLASDRFLGLSLEQARAAVSEAEADFEQNSRYQLTLWGPENGLDYAN 660
Db 599 ELLPALDEVLASDRFLGLSLEQARAAVSEAEADFEQNSRYQLTLWGPENGLDYAN 658
QY 661 KQIAGLVANYTTPRWRLFLDALVDSVAGQIPFQOHOQFDKNVFLQEQAFVLSKORYPSQPR 720
Db 659 KQIAGLVANYTTPRWRLFLDALVDSVAGQIPFQOHOQFDKNVFLQEQAFVLSKORYPSQPR 718
QY 721 GDTVDLAKKIFLKYYP 736
Db 719 GDTVDLAKKIFLKYYP 734

RESULT 3
ID O88325 PRELIMINARY; PRT; 739 AA.
AC O88325;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ALPHA-N-ACETYLGLUCOSAMINIDASE.
GN NAGLU.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RX MEDLINE=20056274; PubMed=10598735;
RA Li H.H., Yu W.H., Rozengurt N., Zhao H.Z., Lyons K.M.,
RA Anagnostaras S., Fangelow M.S., Suzuki K., Vanier M.F., Neufeld E.F.;
RT "Mouse model of Sanfilippo syndrome type B produced by targeted
RT disruption of the gene encoding alpha-N-acetylglucosaminidase.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14505-14510(1999).
DR EMBL; AF003255; AAC26842.1; -.
DR MGD; MGI:1351641; Naglu.
DR InterPro; IPR002086; Aldehyde_dehydr.
```

DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN.1.  
SQ SEQUENCE 739 AA; 82596 MW; 01C71966557CDE27 CRC64;

Query Match 83.8%; Score 3302; DB 11; Length 739;  
Best Local Similarity 82.9%; Pred. No. 2.8e-236;  
Matches 610; Conservative 54; Mismatches 70; Indels 2; Gaps 1;  
QY 1 MEAAVAAAVGVLLLAGAGGAGGADEARAAAVRALVRLGPGPAADFSVVERALAAKP 60  
DB 1 MEAAAGLVILGFLLL--AGGSVGDREARAKAVRELVRLGPGPAANFLVSVERALADES 58  
QY 61 GLDTSYSLGGGGAARVVRGSGTGAAGLHRYLRDFCGCHVAVSGSQRLLRPLPAVPG 120  
DB 59 GLDTSYSLGGGVPVLRGSGTGAAGLHRYLRDFCGCHVAVSGSQRLLRPLPAVPG 118  
QY 121 LTEATPNRYRYQNVCTOSYFVWMDWARWEREDIMWALNGINLALAWSGQEAIWQRVYL 180  
DB 119 LFTETPNRYRYQNVCTOSYFVWMDWARWEREDIMWALNGINLALAWSGQEAIWQRVYL 178  
QY 181 ALGLTQAEINEFFTPAPFLAWGRMGNLHTWDGPPSWHIKOLYLQHRVLDOMBSFGMP 240  
DB 179 ALGLTQSEIDTYTPAPFLAWGRMGNLHTWDGPPSWHIKOLYLQHRVLDOMBSFGMP 238  
QY 241 VLPAPAGHVPAVTRVFPQVNVTKMGSHGFCNSYSFLLAPEDPIPIIGSLFLREL 300  
DB 239 VLPAPAGHVPAVTRVFPQVNVTKMGSHGFCNSYSFLLAPEDPIPIIGSLFLREL 298  
QY 301 KEFGTDHLYGADTFNEMOPPSSEPSYLAATAATVAYEAMTAVDTAVWLLQGLWFOHQPOF 360  
DB 299 KEFGTDHLYGADTFNEMOPPSSEPSYLAATAATVAYEAMTAVDTAVWLLQGLWFOHQPOF 358  
QY 361 WGPQIRAVLGAVRGRLLVLDLFAESQPVYTRTASFGQPFVWMLHNFEGNHLGFGAL 420  
DB 359 WGPQIRAVLGAVRGRLLVLDLFAESQPVYTRTASFGQPFVWMLHNFEGNHLGFGAL 418  
QY 421 EAVNGGPEAARLFNSTWMTGMAPEGISQNEVYVSLMAELGWRKDPVPLAAWVTSFAA 480  
DB 419 EDVNRGQPAARLFNSTWMTGMAPEGISQNEVYVSLMAELGWRKDPVPLAAWVTSFAA 478  
QY 481 RRYGVSHPDAGAAWRLLRLSYNCSGACRGNHNSPLVRRPSLQMTS IWNRSDFEAW 540  
DB 479 RRYGVSHPDAGAAWRLLRLSYNCSGACRGNHNSPLVRRPSLQMTS IWNRSDFEAW 538  
QY 541 RLILTSAPLTSAPFYDLDLTDTRQAVQELVSLYEEARSAYLSKELASLLRAGGVLAY 600  
DB 539 RLILTSAPLTSAPFYDLDLTDTRQAVQELVSLYEEARSAYLSKELASLLRAGGVLAY 598  
QY 601 ELLPALDEVLASDSRFLGSLWLEQARAAVSEAEADFEQNSRYLTWLGPEGNILDYAN 660  
DB 599 KLLPTLDELLASSHFLGTLWLDQARAAVSEAEADFEQNSRYLTWLGPEGNILDYAN 658  
QY 661 KQLAGLVANYTTPRWRLFLDALVDSVAQGIPIFQHQDKNVFQLEAFVLSKORYPSOPR 720  
DB 659 KQLAGLVANYTTPRWRLFLDALVDSVAQGIPIFQHQDKNVFQLEAFVLSKORYPSOPR 718  
QY 721 GDTVDLAKKIFLKYYP 736  
DB 719 GDTVDLAKKIFLKYYP 734

RESULT 4  
Q90276 PRELIMINARY; PRT; 753 AA.  
AC Q90276;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE LYCOSOMAL ALPHA-N-ACETYL GLUCOSAMINIDASE.  
OS Dromaius novae-hollandiae (Emu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Palaeognathae; Casuariiformes; Dromaiidae;  
OC Dromaius.

OX NCBL\_TaxID=8790;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21309063; PubMed=11414757;  
RA Aronovich E.L., Johnston J.M., Wang P., Giger U., Whitley C.B.;  
RT "Molecular Basis of Mucopolysaccharidosis Type IIIB in Emu (Dromaius  
novae-hollandiae): An Avian Model of Sanfilippo Syndrome Type B.";  
RL Genomics 74:299-305(2001).  
DR EMBL; AF331668; AAK73654.1; -.  
SQ SEQUENCE 753 AA; 84013 MW; C4B5332A6775EC01 CRC64;

Query Match 66.3%; Score 2613.5; DB 13; Length 753;  
Best Local Similarity 64.7%; Pred. No. 3.3e-185;  
Matches 480; Conservative 105; Mismatches 144; Indels 13; Gaps 6;  
QY 4 VAVAAAVGVLLLAGAGG-----AAGDEAREAAAVRALVRLGPGPAADFSVVERAL 56  
DB 5 VRLTALAAMAAARAASFPPLPPVRAGAEADARQAAVRLARRLLGPRAA--VALSVDSGL 63  
QY 57 AAKPGDLYSLGG--GGAARVVRGSGTGAAGLHRYLRDFCGCHVAVSGSQRLLRPL 114  
DB 64 AA--GGDLYRVHSPGAAVAVAVAGSGVAAAGLHRYLRDFCGCHLSWSGRQLRLDPL 122  
QY 115 PAVPGELTEATPNRYRYQNVCTOSYFVWMDWARWEREDIMWALNGINLALAWSGQEA 174  
DB 123 PRVPAEIRATAPGRFRYQNVCTOSYFVWMDWARWEREDIMWALNGINLALAFQGEAV 182  
QY 175 WQRYLALGALQAEINEFFTPAPFLAWGRMGNLHTWDGPPSWHIKOLYLQHRVLDOMR 234  
DB 183 WQRYLALGALQAEINEFFTPAPFLAWGRMGNLHTWDGPPSWHIKOLYLQHRVLDOMR 242  
QY 235 SFGMTVPVLPAPAGHVPAVTRVFPQVNVTKMGSHGFCNSYSFLLAPEDPIPIIGSL 294  
DB 243 SLGMITVLPAPAGHVPAVTRVFPQVNVTKMGSHGFCNSYSFLLAPEDPIPIIGSL 302  
QY 295 FLRELKEFGTDHLYGADTFNEMOPPSSEPSYLAATAATVAYEAMTAVDTAVWLLQGLW 354  
DB 303 FLRELKEFGTDHLYGADTFNEMOPPSSEPSYLAATAATVAYEAMTAVDTAVWLLQGLW 362  
QY 355 QHQPQWFGPAQIRAVLGAVRGRLLVLDLFAESQPVYTRTASFGQPFVWMLHNFEGNH 414  
DB 363 QHQPQWFGPAQIRAVLGAVRGRLLVLDLFAESQPVYTRTASFGQPFVWMLHNFEGNH 422  
QY 415 GLFGALVAVNGGPEAARLFNSTWMTGMAPEGISQNEVYVSLMAELGWRKDPVPLAAW 474  
DB 423 GLFGALVAVNGGPEAARLFNSTWMTGMAPEGISQNEVYVSLMAELGWRKDPVPLAAW 481  
QY 475 VTSFAARRYGVSHPDAGAAWRLLRLSYNCSGACRGNHNSPLVRRPSLQMTS IWNRS 534  
DB 482 VARYAERRYGAPNAAASAAWQLLLRSVYNTG--VCYNHNSPLVRRPSLQMTS IWNRS 540  
QY 535 DYFEARLLTSAPLTSAPFYDLDLTDTRQAVQELVSLYEEARSAYLSKELASLLRA 594  
DB 541 DYFEARLLTSAPLTSAPFYDLDLTDTRQAVQELVSLYEEARSAYLSKELASLLRA 600  
QY 595 GGVLAYELLPALDEVLASDSRFLGSLWLEQARAAVSEAEADFEQNSRYLTWLGPEGN 654  
DB 601 GGVLAYELLPALDEVLASDSRFLGSLWLEQARAAVSEAEADFEQNSRYLTWLGPEGN 660  
QY 655 ILDYANKQLAGLVANYTTPRWRLFLDALVDSVAQGIPIFQHQDKNVFQLEAFVLSKOR 714  
DB 661 ILDYANKQLAGLVANYTTPRWRLFLDALVDSVAQGIPIFQHQDKNVFQLEAFVLSKOR 720  
QY 715 YPSQPRGDPVLDLAKKIFLKYYP 736  
DB 721 YPTAPVGDFTLEISKIFLKYYP 742

RESULT 5  
Q90275 PRELIMINARY; PRT; 753 AA.  
ID Q90275  
AC Q90275;

|                  |   |  |
|------------------|---|--|
| DT               | 01-DEC-2001 (TReMBLrel. 19, Created)                                    |  |
| DT               | 01-DEC-2001 (TReMBLrel. 19, Last annotation update)                     |  |
| DT               | 01-DEC-2001 (TReMBLrel. 19, Last annotation update)                     |  |
| DE               | LYSOSOMAL ALPHA-N-ACETYL GLUCOSAMINIDASE.                               |  |
| OS               | Dromaius novaehollandiae (Emu).   |  |
| OC               | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;       |  |
| OC               | Archosauria; Aves; Palaeognathae; Casuariiformes; Dromalidae;           |  |
| OC               | Dromaius.   |  |
| NCBI_TaxID=8790; |   |  |
| ON               | [1]   |  |
| ON               | SEQUENCE FROM N.A.  |  |
| RP               | MEDLINE=21309063; PubMed=11414757;                                      |  |
| RP               | Aronovich E.L., Johnston J.M., Wang P., Giger U., Whitley C.B.;         |  |
| RA               | "Molecular Basis of Mucopolysaccharidosis Type IIIB in Emu (Dromaius    |  |
| RT               | novaehollandiae): An Avian Model of Sanfilippo Syndrome Type B.;"       |  |
| RL               | Genomics 74:299-305(2001).  |  |
| DR               | EMBL; AF331669; AAK73655.1; -.  |  |
| SQL              | SEQUENCE 753 AA; 84008 MW; 3B0445E30B48AA85 CRC64;                      |  |
|                  | Query Match 66.3%; Score 2611.5; DB 13; Length 753;                     |  |
|                  | Best Local Similarity 64.7%; Pred. No. 4.6e-185;                        |  |
|                  | Matches 480; Conservative 104; Mismatches 145; Indels 13; Gaps          |  |
| QY               | 4 VAVAAAVGVLLLAGAGG-----AAGDEAREAAAVRALVARLLGGPAADFSVERAL 56            |  |
| DB               | 5 VRLTALAAWMAAARAASPLPXVRACGAEDARQEAARALARRLLGPRAAA-VALSVDSGL 63        |  |
| QY               | 57 AAKPGLDITYSLGG--GGAARVRVSGTGVAAAGLHRYLRDFCGCHVAGSGLRPLRPRL 114       |  |
| DB               | 64 -AA-GGLDTYRVHSPGCAAVAVAGSSGVAAAGLHRYLRDLCGCHLSGSQLRPLPOPL 122        |  |
| QY               | 115 PAVGELTEATPNRYRYQNVCTQSYSFVMDWARWEREDIMWALINGINALAWSGQBAI 174       |  |
| DB               | 123 PRVPAETRTATPGFRYYQNVCTQSYSFANWDWARWEEEDIMWALSGINALAFAGQAV 182       |  |
| QY               | 175 WQRYVLALGQTQAEINIEFFTGPAFLAWGRGNLIHTWDGPLPPSWHIKQYLQHRVLDOMR 234    |  |
| DB               | 183 WQRYVLSLGNLQSEIDYETFGPAFLAWNRMGNLHGWAGLPRAWHLKQLYQYRVLEMR 242       |  |
| QY               | 235 SFGMTVPLPAFAGHVPAAVTRVPPQVNVTKMGSWGHNCSYSCSFLPADPEDIFFPLISL 294     |  |
| DB               | 243 SLGMITVLPFAFGHVGQVLRAPRVNATRLGWSHFDTCTYCTYLLDPEDPMFQVIGTL 302       |  |
| QY               | 295 FLRELKEFGTDHYTGADTFENQPPSPSEPSYLAATAATVYEAMTAVDTAEVWLQGWLF 354      |  |
| DB               | 303 FLKELKEFGTDHYTSADTFENMNPLSSDPAYLSRVSSAVFRSMTCADPKAVLMQGWLF 362      |  |
| QY               | 355 QHQPFQWGPQAIRAVLGAVPRGLLVLDLFAESQPVYTRTASFGQPTFWCHLHFNFGNH 414      |  |
| DB               | 363 QHQPDFQWQPAQVRAALLHGVPLGRMVLVDLFAESRPVYQWTESTFYGQPTFWCHLHFNFGNH 422 |  |
| QY               | 415 GLFGALSAVNGGPPAARLFPNSTMVGCTGMAPEGISQNEVYVLSMALBGRKRDVPPDLAAW 474   |  |
| DB               | 423 GLGCTVEAINHGPPAARFPNSTMVGCTGLYPEGIEQNDVYVELMNLGNRQBP-LDPSW 481      |  |
| QY               | 475 VTSFAARYGVSHDPDGAANRLLRSVYNGSGEACGRNRSPLVRPRLQMTSTIWNRS 534         |  |
| DB               | 482 VARYAERYGAPNAAAAASAWKLLRSVYNGCTG-VCVNHNRSLVRPRLSRMDTEVWYKNS 540     |  |
| QY               | 535 DYFEAWRLLTSAFSLATSFAFRYDLDLTRQAVQELVSLVYEARSAYLSKELASLLRA 594       |  |
| DB               | 541 DYENAWRLLSAGALEGSSPTFGYDLADVTRQAQQLVSEYVLSIRQAQSRSLPELLTA 600       |  |
| QY               | 595 GGVLAYELLALPDVLEASDRSLGLSGLEQARAAAVSEAEADYVENSRYQLTLWGPEGN 654      |  |
| DB               | 601 GGVLVYDLLPELDGLSSHLRFLGLRWLESARAVATSDREAEQVELNARNQVTLWGPNGN 720     |  |
| QY               | 655 ILDYANKQLAGLVANYTPWRRLFEALVDSVAGQIFPQOHQFDKNVFOLEQAQVLSKOR 714      |  |
| DB               | 661 ILDYANKQLGGLVDIYGVNRSLFVSALVSLNCGSPFHQDQFNQAVQVQVFERGFYKNR 736      |  |
| QY               | 715 YPSOPRGDVTDLAKKIFLKYPY 736  |  |





|        |   |  |         |
|--------|---|--|---------|
| QY     | 248   | HVPEAVTRVFPQVNTVTKMGSWGHNFCNSYSCSFLAPEDPIFPFIIGSLRELRIKFEFGTDH | 307     |
| Db     | 267   | HVPRALKRLNPSTFMEVQRWQNFDPDYCCGLFVEPTENLKEIASRELFHLNIITKYGSNH   | 326     |
| QY     | 308   | IYGADTFNEMQPPSSPSYLAATTAIVYEAAMTAVTEAVWLLQGLWFQHQFQFNGPAQIR    | 367     |
| Db     | 327   | IFTCDFNELEPPVAKPEYMRSTAALVIESMRGIDPQAIWLQGMFMVKNP-FWTTDMAE     | 385     |
| QY     | 368   | AVLGAVPRGLLVDLFAESQPVYTRTASFQGPFTWCMHNFNGHGLFGALEAVNGCP        | 427     |
| Db     | 386   | AFTAAPRGRILVDLQSQFFQYELTRSFQPIWCMHNFNGTGLMFGSAKLINGSI          | 445     |
| QY     | 428   | EAARLPNSTWGTGMAPEGISQNEVYVYSLMAELGWRKDPVDPDLAANTVSTAARYGVSH    | 487     |
| Db     | 446   | EEARRLPNSSLVGTITPEGIGQYVMSFTLGRGWSNTSL-DLDSWETNFESHSGYVKD      | 504     |
| QY     | 488   | PDGAAWRLLRVYVNCSG-EACRGHNRSPLVRRPSLOWNTSIWVNRSDVFEAWRLLLTS     | 546     |
| Db     | 505   | ERLEQAWLLKNVSYFRGLQKNRG--QVYVTRRPSFNQEPETWYNASAVIDAWHLLTTF     | 562     |
| QY     | 547   | APSLATS---PAFRYDLDLTRQAAQVELSVLYEARSAYLSKELASLLRAGGVLAYEL      | 602     |
| Db     | 563   | RAIPLDNRNVEIYEDLVDITRFLQISADQLYINLRSAYRKRQVSRF----EFLSVKL      | 618     |
| QY     | 603   | LPALDE---VLASDRFLGSLGWLEQARAARAAVSEAEADFYEQNSRQTLTWGPEGNILDYA  | 659     |
| Db     | 619   | LKLFDDMELTASRNFLLGNWLQQAKQAPNTGQORNEEFNARNQITAWGPDGQLLDYA      | 678     |
| QY     | 660   | NKQLAGLVAVNYPWRFLFEALVDSVAOCIPQHQHFDKNT-FQLEQAFVLSKORYPSQ      | 718     |
| Db     | 679   | CKQWSGLVSDYYPWRFLFEDYTVLHAGRPNGTAFKLKYSHEIETLFPFNKDDVYPVT      | 738     |
| QY     | 719   | PRGDTVDLAKKIF 731  |         |
| Db     | 739   | PVGNTWLISQDIF 751  |         |
| RESULT | 9   |  |         |
| Q9VLL5 |   |  |         |
| ID     | Q9VLL5  | PRELIMINARY;   | 690 AA. |
| AC     | Q9VLL5;   |  |         |
| DT     | 01-MAY-2000 (TREMBlrel. 13, Created)                                  |  |         |
| DT     | 01-MAY-2000 (TREMBlrel. 13, Last sequence update)                     |  |         |
| DT     | 01-JUN-2001 (TREMBlrel. 17, Last annotation update)                   |  |         |
| DE     | CG13397 PROTEIN.  |  |         |
| GN     | ESTS:172F5T OR CG13397.   |  |         |
| OS     | Drosophila melanogaster (Fruit fly).                                  |  |         |
| OC     | Eukaryota: Metazoa: Arthropoda; Tracheata: Hexapoda; Insecta:         |  |         |
| OC     | Pterygota: Neoptera: Endopterygota; Diptera; Brachycera; Muscomorpha; |  |         |
| OC     | Ephydroidea; Drosophilidae; Drosophila.                               |  |         |
| OX     | NCBI_TaxID=7227;  |  |         |
| RP     | [1]   |  |         |
| RC     | SEQUENCE FROM N.A.  |  |         |
| RC     | STRAIN=BERKELEY.  |  |         |
| RX     | MEDLINE=MD196006; PubMed=10731132;                                    |  |         |
| RA     | Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,       |  |         |
| RA     | Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,       |  |         |
| RA     | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,   |  |         |
| RA     | Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,         |  |         |
| RA     | Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  |  |         |
| RA     | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., |  |         |
| RA     | Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., |  |         |
| RA     | Ball R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,     |  |         |
| RA     | Besoon K.Y., Bencs P.V., Bernan B.P., Bhandari D., Bolshakov S.,      |  |         |
| RA     | Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,        |  |         |
| RA     | Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., |  |         |
| RA     | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,         |  |         |
| RA     | de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,     |  |         |
| RA     | Doudou K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P. |  |         |
| RA     | Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W. |  |         |
| RA     | Foster C., Gabrielian A.E., Garg N.C., Gelbart W.M., Glasser K.,      |  |         |
| RA     | Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,          |  |         |

```
QY 723 TVDLAKKIF 731
D 1 : : : :
Db 655 TWLISQDIF 663

RESULT 10
Q9AAQ6
ID Q9AAQ6 PRELIMINARY; PRT; 770 AA.
AC Q9AAQ6
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ALPHA-N-ACETYLGALUCOSAMINIDASE.
GN CC0540.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; A5005727; AAK22527.1; -.
DR TIGR; CC0540; -.
KW Complete proteome.
SQ SEQUENCE 770 AA; 84287 MW; 94A694F41C18DB20 CRC64;

Query Match 26.9%; Score 1060.5; DB 16; Length 770;
Best Local Similarity 33.0%; Pred. No. 5e-70;
Matches 260; Conservative 132; Mismatches 302; Indels 93; Gaps 19;

QY 2 EAVAVAAVGVLLAGAGAGAGDEAREAAVVALVRLGPGPAADFVSVERALAAKPG 61
D 1 : : : :
Db 11 QAISLAAAFVAPALAAAGSTG-VAARASLRKLFGRRLAG-----AHLTVTPG 59

QY 62 LDT--YSLGGGAARVVRGSGTGAAGLHRYLRDFCGCHVAVSGSOLRLPRPLPAVPG 119
D 1 : : : :
Db 60 AERSWYAIGKGA-ISISGSPVALVRGAYAHRLQAGLAHVSWEGRVQVAGVAPAG 118

QY 120 ELTEATPNRYQNVCTQSYFVVDWAWREIDMMLNGINLALAWSQEAIWQVRY 179
D 1 : : : :
Db 119 ARVE-TPFRRAYLNTCTGYTTPWGMGRWTRIDWMAAHGIDMPLAMEGQEVVRLW 177

QY 180 LALGLTQAEINEFFTPAFLANGMNLHTDGLPPSWHIKQLYLOHRLVLDOMRSGMT 239
D 1 : : : :
Db 178 REFLSEALADYFSGPAFTPHRMGNIEGYKAPLTAWIDKKDLQVKILGRMSLGMT 237

QY 240 PVLPAFAGHVAEVRVFPQVNVTKMGSGHFNCSYCSFLIAPEDPIFFPIIGSLFLREL 299
D 1 : : : :
Db 238 PLPAFGGVVPKFAEKPKARIYRMWPWEGFHETI-----WLDPAFLFAKIAARFLALY 293

QY 300 IKEFGTDHYGADTNEWQPP-----SSEPSY-----LA 328
D 1 : : : :
Db 294 TETFGAGTYLADSFNEMLPPINADGADARDAAYGTANTAVTKVEVDPAKKAORLA 353

QY 329 AATTAYEAMTAVDTEAVMLLQGLWLFQHPQFQWGAQVRAVLGAVPRGLVLVDLFAEQ 388
D 1 : : : :
Db 354 AVGKAIYDSIRCTRDVAWVWQGLFGADSHFWDPAALSAVLSLVPDDKMLILDIGNRY 413

QY 389 P-VYTRTASFQOPTICMLHNFQGNHGLFGAL-----EAVNGGPEAARLFPNSTMVG 440
D 1 : : : :
Db 414 PNWKNAKAFGKWPYGVYVHNYGSGNPVYDGLGYRQDIPAIANPDAGL-----AG 467
```

```
QY 441 TGMAGEISQNEVYVSLMAELGWRKDPVDPDLAAWVTSFAARRYGVSHPDAGAAWRLLS 500
D 1 : : : :
Db 468 FGMFPEGLHNNSIVYEAVDLAWSEQASP-ATWLTRYARARYKTSPLDALGOLVPA 526

QY 501 VYNCSEACRGNRSP-----LVRRPSLQWNTSIWY--NRSDVFEAWRLLLTSAPSLS 554
D 1 : : : :
Db 527 AFSTRYWSRPMWMSKAGAYLFFKRPRTATYGFDFPHQGDRAKLEAAVKALTALPTYGQP 586

QY 555 AFRYDLDLTROAVOELVSLYYEARSAYLSKELASLLRAGGVLAVELLPALDEVLASD 614
D 1 : : : :
Db 587 LFVLDLTDATRLHATMKIDDLQVAAVYRGDTA-----AGDAARVEI-----EALASI 637

QY 615 RELLG-----SWLEQARAAVSEAEADYEQNSRYQLTLWPGEGNILDYANKQLAGLV 667
D 1 : : : :
Db 638 DKLLGVPQDTLATWIDEARAYGDTADAAAYVANAKAQTWINGEGNLDYASKAWQGLY 697

QY 668 ANYYTPRWLFLEALVDSVAQGIPIQOHQFDKNVQLQEAQFVLSKQRY-----PSOPRGDT 723
D 1 : : : :
Db 698 KSFYLPWRSRFLDAL--KAAGTGTDEVTVTGGVAVERAWVEAYVRRKPADPIGEI 755

QY 724 VDLAKKI 730
D 1 : :
Db 756 KTLIARI 762

RESULT 11
Q9NAP6
ID Q9NAP6 PRELIMINARY; PRT; 575 AA.
AC Q9NAP6
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE K09E4.4 PROTEIN.
GN K09E4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z83234; CAB70170.1; -.
SQ SEQUENCE 575 AA; 68849 MW; FE7216DB39F4D93C CRC64;

Query Match 24.6%; Score 970; DB 5; Length 575;
Best Local Similarity 35.7%; Pred. No. 1.7e-63;
Matches 210; Conservative 104; Mismatches 248; Indels 26; Gaps 10;

QY 135 VCTQSYFVVDWAWREIDMMLNGINLALAWSQEAIWQVRYLALGLTQAEINEFF 194
D 1 : : : :
Db 1 MCTSYFSAWVEWQWERFIWIALNGFNTVIMPLGOEIIWRDIFMGLGVQVDELDSYFT 60

QY 195 GPAFLANGMNLHTDGLPPSWHIKQLYLOHRLVLDOMRSGMTVPLPAFAGHVAEVR 254
D 1 : : : :
Db 61 SQAVLAWHRMGNLKAYGGGLSDAQLNDHNLAKRIIDRLLELGTPIPLTFAGVDPHLE 120

QY 255 RVFPQVNVTKMGSGHFNCSYCSFLIAPEDPIFFPIIGSLFLRELKEFGTD--HIYGA 312
D 1 : : : :
Db 121 TLFPASKENRLPRNNFTSETSCMLSVSPEDPLFKIGSTFLRHOKKMFQGDVNMYSAD 180

QY 313 TFENMQPSS---EPSYLAATTAATVYEAATVDEAVWLLQGLWLFQHPQFQWGAQVRAV 369
D 1 : : : :
Db 181 PFNEILPSESAKFADKAEKVFVQTAQAIMNSCKKVDKNCVWVLLQSWSFYTDQ--WPAWAKSF 238
```

```
QY 370 LGAVPRGLLVLDLFAESOPVYTRTASFOGQPFIMCMLHNFEGNHGLFGALEAVNGGPEA 429
Db 239 LSAIPVGNLLIIDLVAEVPVQWMTSSFOGHFVWCLLHNFEGSRELKGNLQKIDKGQL 298
QY 430 ARLFNPSTMVGTGMAPEGISQNEVYVSLMAELGWRKDPVPDLAAWVTSFAARYGVSHPD 489
Db 299 ALMKAGSNLVGAGLSNEADQNVVYVQFMIDRWMSPEPLP-LNNMLKAYSESRYSDAFKV 357
QY 490 AGAAMPLLLRSVYNGSGEACRGHNSPLVRFSLOMNTSIWYNSRDVFEAWRLLLTS-AP 548
Db 358 AQKFWTLLAGTYNQPEKMGWTPRESVFLYHREFGCRKIEYFPVEETFSRFRLELLPALVH 417
QY 549 SLATSPAFRYDLDLDTROAVQELVSLVYEEARSAYLSKELASLL-----RAGGVLAYELLP 604
Db 418 VLGEHPLFREDLNDVNMETQ-----FEMGNEAALSSEAFMEDKQGVGA--SCLEML 469
QY 605 ALDEVLASDRFLGSLWLEQARAAVSEAEADFYENSRVQLTLWGPEGNILDYANKQLA 664
Db 470 EMFOKLESYNSRDVROWIENAKSIAPTSEERQVFPVTTAGDILTVMGPTGNLDYAHREWA 529
QY 665 GLVANYTTPRWRLFLLEALYDSDVAOGIPFOQHOQDFKNVFO-LEQAFVLS 711
Db 530 GLMSGYGRRWQYF-----CDWILEHDFNHTFESISVFRDVERPFSIS 573

RESULT 12
Q93TW7 PRELIMINARY; PRT; 1862 AA.
AC Q93TW7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MXAE.
GN MXAE.
OS Stigmatella aurantiaca.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
OX NCBI_TaxID=41;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21110452; PubMed=11182319;
RA Sliakowski B., Nordsiek G., Kunze B., Blocker H., Muller R.;
RT "Novel features in a combined polyketide synthase/non-ribosomal
RT peptide synthetase: the myxalamid biosynthetic gene cluster of the
RT myxobacterium Stigmatella aurantiaca Sgal5.";
RL Chem. Biol. 8:59-69(2001).
DR EMBL: AF319998; AAK57189.1; -.
SQ SEQUENCE 1862 AA; 200005 MW; E809DCEC6A320144 CRC64;

Query Match 3.6%; Score 143.5; DB 2; Length 1862;
Best Local Similarity 22.3%; Pred. No. 0.19;
Matches 180; Conservative 79; Mismatches 289; Indels 259; Gaps 37;

QY 3 AVAAVAVGVLLLAGAGGAAGDEAREAAVVALVRLGPGPAADFSVVERALAAKPG 62
Db 962 AVPASALVELGLSAAAEALGAGARELSDVEFERALVLTADRLVQVHLSPASGGQHV 1021
QY 63 DTVSLGGGAA-----RVRVRSSTGVAAAAGLHRYLDRFCGCHVAVSGSLRLPR- 112
Db 1022 HHSRAVGGTSPGAGWVRHRCQIRAAITPSGSPVLDVRSRCAQHVQGAAYVEQLERC 1081
QY 113 -----PLPAV-----PGEUTEATPNRYRYQWCTQSYFVWMDWARERE----IDWM 157
Db 1082 NVQYEAPLRTLGEARRRPGALGAVALSPEVIQ-----ESARYQLPALD-A 1128
QY 158 ALNGINLALASGOEAIWQRYLALGLTQAEINEFFTGPAFLAWG-----RMG 205
Db 1129 GLQTLALALAAESGEAV---LFMPLSTGSLECVQ---GRADVKAHVSIANAATSPEDRM 1182
QY 206 NLHTWDG-----PLPPSHWIKQLYLOH---RVLDQMR-----FGMTPVLPAFAGHVP 250
```

```
Db 1183 TLELDGEGRRVAVARGVRLRRVAAERLLEVLGSEARSOEHFDVWEPRAVTAQPGPAD 1242
QY 251 EAVTRVFPQVNVTKMSGHF-----NCYSYSCSFLLAPE-----DPIFFPIIG 292
Db 1243 WLV-----FLDRGGWGHTALVEEIGRQGPQCVTVTAGETTFQRQDARFRVNVKRPEDM 1294
QY 293 SLFLRELII-----KEFGVDHYIGADTFNEMQPPSPSEPSYLAANTATVYEA- 338
Db 1295 ERLRELPAIPAGHEGRAVYVWGLDAVLDEQGTGPES--VAALHLVKALMGSPARALW 1352
QY 339 -----TAVDTAVWLLQWLF-----QHOPQEWGPAQIRAVLGA VPRGRLLVLIDL 383
Db 1353 VYTRGAQVTGCTERVSLAQAPLWGMGRSVLEQPCWGG-----LIDL 1396
QY 384 FAESOPVYTRTASFOGQPFIMCMLHNFEGNHGLFGALE-----AVNGGPEAR 431
Db 1397 APER--VAEETAALVRE-----ISAFGGDGEDQIALREKRSVPRIARGRVNAPAEPLR 1448
QY 432 LPPNSTMVTGTHAPEGISQNEVYVSLMAELGWRKDPVPDLAAWVTSFAARYGVSHPDAG 491
Db 1449 LRPDGAYLVTTGG-----LGLGLIK-----VARWLVARGARHL----- 1480
QY 492 AAWRLLLRSVYNGSGEACRGHNSPLVRFPRLSQMNTSIWYNSRDVFEAWRLLLTSAPLS 551
Db 1481 ----VLLGRSGASGAGDAASARREGI-----ESLRALGASVTTLA 1517
QY 552 TSPAFRYDLDLDTROAVQELVSLVYEEARSAYLSKELASLLRAGGVLAYELLPALDEVLA 611
Db 1518 VDVAOREKLAALLREAAATL-----PPLRGVIHAAALLTESLNEMD--LA 1561
QY 612 SDSREL----LGSWL--EQARAAVSEAEADFYENSRVQLTLWGPEGNILDYANKQLAG 665
Db 1562 AMTAMMRPKVGLSWLHVETR-----EALDFFVFMESSTS-TLWASG-LAHYA----AG 1610
QY 666 LVANYTTPRWRLFLLEALV-DSVAOGIP 691
Db 1611 -----NQFLEALAHHRRAQGLP 1627

RESULT 13
Q96W93 PRELIMINARY; PRT; 444 AA.
AC Q96W93;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TR18.
GN TR18.
OS Gibberella zeae (Fusarium graminearum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=88-1;
RX MEDLINE=21318682; PubMed=11425709;
RA Lee T., Oh D.W., Kim H.S., Lee J., Kim Y.H., Yun S.H., Lee Y.W.;
RT "Identification of Deoxyvalenol- and Nivalenol-Producing Chemotypes
RT of Gibberella zeae by Using PCR.";
RL Appl. Environ. Microbiol. 67:2966-2972(2001).
DR EMBL: AF336365; AAK53574.1; -.
SQ SEQUENCE 444 AA; 47659 MW; BC5208C72484C701 CRC64;
```

```
Query Match 3.2%; Score 128; DB 3; Length 444;
Best Local Similarity 20.3%; Pred. No. 0.32;
Matches 97; Conservative 54; Mismatches 160; Indels 166; Gaps 24;

QY 157 MALNGINLALA-WSGOEAIWQRYLALGLTQAEINEFFTGPAFLAWGRMGNLHTWDGPLP 215
Db 1 MALDRLLFLLSLWLG-----FVGAQAALSE-----PLP 29
QY 216 PS---WHIKOLYLOHRLVDQMRSGMTVPVLPAPAGHVPEAVTRFPQV-NVTXMSGNGHF 271
```



```
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwytzer M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=JURA;
RX MEDLINE=971164286; PubMed=9010999;
RA Schwytzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA LaBoissiere S., Misra V., Vleck C., Paces V.;
RT "Gene contents in a 31-kb segment at the left genome end of bovine
RL herpesvirus-1.";
RN Vet. Microbiol. 53:67-77(1996).
RP [5]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwytzer M., Vleck C., Lowery D.E., Bello L.J., Meyer G., Misra V.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwytzer M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z78205; CAB01605.1; -.
DR EMBL; AJ004801; CA006097.1; -.
SQ SEQUENCE 3247 AA; 332190 MW; 3AEAA72F8F001F6A CRC64;

Query Match          3.1%; Score 122.5; DB 12; Length 3247;
Best Local Similarity 22.9%; Pred. No. 15;
Matches 192; Conservative 68; Mismatches 297; Indels 283; Gaps 45;

QY 1 MEAVAVAAVGVLLLAGAGGAAGDEAREA--AAVRAL--VAR---LIGPGPAADFSVSVE 53
DB 873 MSAARAAARAAGA--AAEAPGAGAEADAAEAARVGGADDVLRQLDALATCAAADREARA- 929
QY 54 RALAAKPLGDTYSLGGGGAARVRVGTGVAANAAGLHRY-----LRDFCGCHVAMSGSQ 107
DB 930 RVAAARAGLEAQAAL---AARRRY-AELGERRAELYRLDALLPLQGFAG----- 975
QY 108 LRLPRPLPAVPGELTEATPNRYRYQNVCTQSYSFVMDWAREREIDMMAINGINLALA 167
DB 976 -----LRAAPGAL-----ERLCADAGAQSAEDCARFLOEAPPVRAAGV----- 1013
QY 168 WSGQEAIRQRYVYALGLTQAEINEFFTGPAFLAWGRMGNLHTWDGPLPPSHIKQLYLQH 227
DB 1014 --QDRLWOLF-----GRYREALNPSSALAPGALAGL----GPAAFA----- 1048
QY 228 RVLDQMRSGMTVPVLPD-FAGH----VPEAVTRVFPQVNVTKMSGWHGNCYSYCSFLLAP 283
DB 1049 -VLGTALQAMGPAARAFEFGEHAERVAARAAVAARAAE-----P 1084
QY 284 EDPIFFPIIGSLFLRELILKEF-GTDHT-----YGDATFENEMQPP 320
DB 1085 AAPAAVAAAASALREAAEALRAGGAHVPPPEPAFLDALQAQAAARREAOQTGARRLGAALAA 1144
QY 321 SSEPS-YLAAATTAAYEAMT-----AVDTEAWLLQGMFLQHQPOFGWPAQIRAYVLGAVPR 375
DB 1145 VSEASGALASAASALKNAWTAGADAETA-----AAAAADAALGAEE 1188
QY 376 GRLLVLDFLFAESQPVYTTASFGQPFITWCMHNFNGHGLFGALEAVNGGPEAARLPFN 435
DB 1189 ----VLRAADEA-----AVAAGGE-----DADGGEDASEADAGDAGDAPPAGRLGPQ 1231
QY 436 STMVCTG-----MAPEGISQNEVYSLMAELGWRKD----- 466
DB 1232 AAKLHADLRKARALARRRAEEVRRLRAEERRAEGAAARQOEERWRDLRAALERVEANS 1291
QY 467 --PVPDLAAWYTSFAARRYGVSHPDAGAAWRLLLRSYVNCSG--EACRGHN-RSPL----- 517
DB 1292 AFDAAEALRLRAAAAAARGYD-PRPLARQADRALEASQAQATAATEAVLAFNPHAPENAKI 1350
QY 518 -VRRPSLQWNTSIWYN-----RSDVFEAWRL-----LLTSAPSLATSPAF 556
```

```
DB 1351 DVPPPLAQLRGIAWMDAFSLAAPALSALFPAADVGQLTRLMHIATGLLTFAASGGGEPRY 1410
QY 557 RYDLL-----DLTROAVQELVS--LYEEARSAYLSKELASLLRAGGVLAYELLPALDEV 609
DB 1411 -YDAVTYLEGDLA--AVPLAKYVAFYRRGHAFF-EERARL-----GALRADVLQAG-- 1460
QY 610 LASDSRFLGSLWLEQARAAAVSEAEADFYEQNSRYQLTLWGPEGNILDYANKQLAGLVAN 669
DB 1461 -----GRAAEISRAR-----RGNVLRSPAERRA-LEAG 1487
QY 670 YTPRWRLEALVDS-VAQGIPFQOHQDKNVFQLEQAFVLSKQRYPSQPRGDTVDLAK 728
DB 1488 V---RIELPSEALIASALAQLERFDHARFAGSAY---EAQMLSTVRRDAAAAAREALEAAK 1541

Search completed: September 21, 2002, 12:29:08
Job time: 278 sec
```